

OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsura A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338 (2000).
DR EMBL; AP002996; BAB48439.1; --
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0005215; F:transporter activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR GO; GO:0008810; P:transport; IEA.
DR InterPro; IPR000515; BPD transp.
DR InterPro; IPR002933; Peptidase M20.
DR Pfam; PF01546; Peptidase M20; 1.
DR PROSITE; PS00402; BPD TRANS_INN_MEMBER; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 421 AA; 45172 MW; 67640F33C747287E CRC64;

Query Match 31.4%; Score 642.5; DB 16; Length 421;
Best Local Similarity 37.1%; Pred. No. 2.4e-32;
Matches 151; Conservative 68; Mismatches 161; Indels 27; Gaps 6;

QY 10 RIEKEIRLSRSA-BPGVTRLTYPPEHAARETLIAAKKAALSVRDALGNIIGRE 68
DB 19 QIQLFDLKGKGAHSGTGVMRTVYGPWVAADQFAECREAGLAVHSDAVGNWGRLE 78
QY 69 GTDPELPALAVGSHFDSVRNGMGFDGTAGVVCALAEARVLENGYVNRHPEFFIAVERE 128
DB 79 GEEPG-PSIVSGSHIDSQTFGGYDGLGALGAIALVAADALAKQFGKPRTLVLAUCEEE 137
QY 129 GARFSSGML-GGRAIAGLVADRELDLSVDEGVSVRQAATAFGKPGELQAAARSADLR 187
DB 138 GSRFPAASLWGSRAITGCDIPREFDELVDGDSVIGQAMKAVGLDNDLGKAQRD--DIG 195
QY 188 ARIELHIEQGPVLEAAGLVGLVTAINGLRHYRVELAGEANHAGAPPMDARRDPWAGFA- 254
DB 196 TFIELHIEQGPVLEAAGLVGLVTAINGLRHYRVELAGEANHAGAPPMDARRDPWAGFA- 254
QY 248 MYREVNRFVNEIADG-----TVATVGHLTVAPEGGQVGPCEVDFTLDLRSPHEE 296
DB 255 -----EAGGLISTAEWRGPAVTVGQVSEPNLPAIPAKVTFTMIDARHPDP 304
QY 297 SLRVLIDRISVMVGEVASQAGVAADVEFNLSPVQLAPTMVDVAREASALQFTHRDIS 356
DB 305 AVQRLYTLHENLHEVADRGKVKITVVENQVPLICHPEIVTAIKATAEEGIRLSNLS 364
QY 357 SGAGHDSMFIAQVTDVGMVFPVRAGRSHVPEWTFDDDLRKGTEVV 403
DB 365 SGGSHDTQOMRIARAGRIFVRSKGRSHTPEFFSIDDIVDGIKVL 411

RESULT 13

Q8ZBY1
ID Q8ZBY1 PRELIMINARY; PRT; 430 AA.
AC Q8ZBY1
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative amino acid hydrolase.
GN YP03249.
OS Versinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;

OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CO-52 / Biovar Orientalis;
RX MEDLINE=21470413; PubMed=11586360;
RA Parkhill J., Wren B.W., Thomson N.R., Titball R.W., Holden M.T.G.,
RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdano-Tarraga A.M.,
RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
RA Felwell T., Hamlin N., Holroyd S., Jagels K., Karlyshev A.V.,
RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrrell B.G.;
RT "Genome sequence of Versinia pestis, the causative agent of plague";
RL Nature 413:523-527 (2001).
DR EMBL; AJ414156; CAC92484.1; --
DR FIR; AH0394; AH0394.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase M20.
DR InterPro; IPR001261; Pept M20_A_B.
DR Pfam; PF01546; Peptidase M20; 1.
DR PROSITE; PS00758; ARGE DAPE_CPG2_1; 1.
KW Hydrolase; Hypothetical protein; Complete proteome.
SQ SEQUENCE 430 AA; 45510 MW; 9C13B117DED0F466 CRC64;

Query Match 31.4%; Score 642.5; DB 16; Length 430;
Best Local Similarity 36.6%; Pred. No. 2.4e-32;
Matches 145; Conservative 66; Mismatches 172; Indels 13; Gaps 3;

QY 18 LSRFSAEGGVTRLTYPPEHAARETLIAAKKAALSVRDALGNIIGREGTDPALPAI 77
DB 24 LAATSEPEGLTRVLSPEHLRANRQVGEWMOVGMQWQDVTGNIICGYEGRQDPALPAI 83
QY 78 AVGSHFDSVRNGMGFDGTAGVVCALAEARVLENGYVNRH-----PPEFFIAVEEGAR 131
DB 84 LIGSHLDTVRNAGRYDGMGLVLTAEV-----GYLHRRQQLPVAIEVIGFADSEGT 137
QY 132 FSSGMLGRAIAGLVADRELDLSVDEGVSVRQAATAFGKPGELQAAARSADLPAFIE 191
DB 138 FGITLLSGKGVTRGPVFWLNT-TDADGISVAQAMVRAGLDPMIDIGQSARANAFCAYLE 196
QY 192 LHIEQGPVLEAAGLVGLVTAINGLRHYRVELAGEANHAGAPPMDARRDPWAGFA- 251
DB 197 LHIEQGPVLEAAGLVGLVTAINGLRHYRVELAGEANHAGAPPMDARRDPWAGFA- 256
QY 252 VNRVFNVEIADGTVATVGHLTVAPEGGQVGPCEVDFTLDLRSPHEESLRVLIDRISVMVGE 311
DB 257 VEALTAAGQEHVATVGTITCLPGAQVNVIPGQVRLTDIRGPNDRGVNDLLRLLAEEA 316
QY 312 VASQAGVAADVEFNLSPVQLAPTMVDVAREASALQFTHRDISGAGHDSMFIAQVTD 371
DB 317 TATREGITFAAGGFRIRKATACDSALQCCISQVQGRCLALPSGAGHDAIAMAECWP 376
QY 372 VGMVFPVRAGRSHVPEWTFDDDLRKGTEVVLRVM 407
DB 377 VGMVFPVRAGRSHVPEWTFDDDLRKGTEVVLRVM 412

RESULT 14

Q8D176
ID Q8D176 PRELIMINARY; PRT; 431 AA.
AC Q8D176;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Putative N-carbamyl-L-amino acid amidohydrolase.
GN Y0939.
OS Versinia pestis.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Versinia.
OX NCBI_TaxID=632;


```

RESULT 8
Q7WJRI PRELIMINARY; PRT; 420 AA.
AC Q7WJRI
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE N-carbamoyl-L-amino acid amidohydrolase (EC 3.5.1.-).
GN AMAB OR B2432.
OS Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxID=518;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RB50 / ATCC BAA-588;
RX MEDLINE=22827954; PubMed=12910271;
RA Parkhill J., Sebaihia M., Preston A., Murphy L.D., Thomson N.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Harris B., Quail M.A.,
RA Cardeno-Tarraga A.M., Temple L., James K., Basham D., Bason N.,
RA Achtman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitsch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.;
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica.";
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640444; CAE32926.1; -.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 420 AA; 44863 MW; ABAG911CE3AD870B CRC64;

Query Match 33.7%; Score 691; DB 16; Length 420;
Best Local Similarity 37.0%; Pred. No. 2.1e-35;
Matches 153; Conservative 72; Mismatches 185; Indels 4; Gaps 2;

Qy 2 TLQAQAAAEKEIRELSRFSAGPGVTRLTYTPEHAAAARETLIAAMKAAALSVDALG 61
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 4 TFPLNAERLWSRVETLSRMTPDPVTTRAFSPLEFQSROWLAEEFAAGLAFLRDAGG 63

Qy 62 NIIGREGTDPELPATAVGSHFDPSVRNGMGFDGTAGVCALAEARVMLENGYVNRHPFEF 121
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 64 NLVGRPDGAARRPIATGSCHCDTVWAGRFDGIIGVLAGEIVANTLHEOGVTLEHPFEV 123

Qy 122 IAIVEEGARFSSGMGLGGRAIAGLVADRELDSDVGVSVRQAATAFGFKPGEIQAAAR 181
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 124 IDFLSEEPSDYGISCVSRALSQLTP-DMLAARNADGETLAEGLRRIGGDDPALGAPLR 182

Qy 182 SAADLRAFIELHIHQGPILQEIOIEGVTSISVGRALRVAVKGSRDHAGTTPMHLRODA 241
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 183 GPQTAAFVELHIHQGPLRGPLIGVVNTNIGIRRVLIITVHGQPDHAGTMPDIRDA 242

Qy 242 LVPAALMVRNVFNVEIADG---TVATVGHLTVPAGGNQVPGVEDFTDLRSPHESL 298
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 243 LVGAARVIDAAHRRASTQSGKHYPVATTIGRIAMTNPNAPNAVPGVMELLEVRSOAVL 302

Qy 299 RVLIDRISVMVGEVASQAGVAADVDFEFNLSPQLAPTMTWDVAVREAASALOPTHRIDISSG 358
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 303 DAFFEAIMAEVADAALAALSAAMPVSRAPHTDCAPLVMDAVEDAATRLGYASMRLPSG 362

Qy 359 AGHSMPFIQVTDVGMVFVPSRAGRSHVPEEWTFDDLKRKTGVTVLRVNWKALDR 412
| | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 363 AGHDVYMPTGPIGMIFVPCINGRSHCPPEWLSPQLLDGTRVLRYETVLELDR 416


RESULT 9
Q7WALO PRELIMINARY; PRT; 420 AA.
AC Q7WALO
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein PMO160.
GN PMO160.
OS Pasteurella multocida.
```

```

Db 190 IKAYEMHIEQKVLSEHDLISGIVTDIQGPVWDLVTLEGAADHAGATPMDMRKD----A 245
Qy 246 ALMWVENVRFVNEIA--DGTATVGHLTVPAGGQVQVGEVDFTLRLSPHEESLRVLID 303
Db 246 GLAMAEVLLAVSAISKEHGGVGTGKMSIEPGGVNIIIGRACFSVDLRHRRKRRQHMVD 305
Qy 304 RISVMVGEVASQAGVAADVDFEFLNLSVQLAPTMTVDVAVREAAASALQFTHRDISSGAGHDS 363
Db 306 DLHEQVEALCNRGVYTNIDVKKEVEPATCSHEMVGLEDVCTELNIRAMKMPCCAGHDA 365
Qy 364 MEIAQVTDVGMVFPVPSRAGRSHVPEWTFDDLRKGTGVVLRMKAL 410
Db 366 LIMSKLAPITGMIFIRSKQISHSPKESWDAEDCKGKTQVLLTLMKL 412

RESULT 6
Q8QGGS PRELIMINARY; PRT; 409 AA.
AC Q8QGGS;
DT 01-MAR-2003 (TremBLrel. 23, Created)
DT 01-MAR-2003 (TremBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE N-carbamoyl-L-amino acid amidohydrolase.
LN.
OS Geobacillus kaustophilus.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
OX NCBI_TaxID=1462;
RN [1]
RP SEQUENCE FROM N.A.
RT Chien H.R., Hsu W.;
RT "Cloning, expression, and characterization of the thermostable n-
RT carbamoyl-L-amino acid amidohydrolase from Geobacillus kaustophilus
RT cccr1123".
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF425838; AAN31517.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0008508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR001005; Myb_DNA_binding.
DR InterPro; IPR002933; Peptidase M20.
DR Pfam; PF01546; Peptidase M20; 1.
DR PROSITE; PS00037; MYB_1; 1.
KW Hydrolase.
SQ SEQUENCE 409 AA; 44328 MW; 435771D32179A0B1 CRC64;

Query Match 35.2%; Score 720.5; DB 2; Length 409;
Best Local Similarity 37.9%; Pred. No. 2.9e-37;
Matches 151; Conservative 75; Mismatches 169; Indels 3; Gaps 3;

Qy 7 QAARTEKEIRLSRPSAE--GPGVTRLTYPTEHAARRETLIAAKAAALSVRDALGNIIIG 65
Db 3 QGERLWQRLMEIGEVKPKSGGVTRLSFTAEERRAKDLVASYNRAGLVFVEDTAGNLIG 62
Qy 66 REGGTDPELPALAVGSHFDSVRNGMFDGTAGVVCALAAARVLENGYVNRHPFFFIATV 125
Db 63 RREGANPDAPVVLVGLSHDSVYNGGCFDGPLVGLAGVEVVQTMNEHGVVTHPIEVAFT 122
Qy 126 EEEGARFSSGMLGGRAIAGLVADRELDSLVDEGVSVRQAATAFGLKPGELQAAARSAD 185
Db 123 DEEGARFREGMIGSRAMAGTLLPEALECR--DANGISIAEMRQTGLDPDLRQPAARKPT 181
Qy 186 LRAFTLHIEQGPILQEQIEIGVTVTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPA 245
Db 182 VKAYVELHIEQGRVLEEAGLPGVITGIAGLIIVKRFIEEGKAHAGATPMSLRDPMAAA 241
Qy 246 ALMWVENVRFVNEIADGTATVGHLTVPAGGQVQVGEVDFTLRLSPHEESLRVLIDRI 305
Db 242 AQIITVIEEARRTGT--TVGTGQLHVYEGGINVPERVEFVLDLRLKAEVRDQWMDI 300
Qy 306 SVMVGEVASQAGVAADVDFEFLNLSVQLAPTMTVDVAVREAAASALQFTHRDISSGAGHDSMF 365
```

```

Db 301 ASRAETIAKERNVRLTTERLQEMAPVLCSEVVVQAARACKQLGYPPWLPSPGAHDGVQ 360
Qy 366 IAQVTDVGMVFPVPSRAGRSHVPEWTFDDLRKGTGVV 403
Db 361 LAPICPIGMIFVRSDGVSHSPABWSTKEDCAVGAEVL 398

RESULT 7
Q89H48 PRELIMINARY; PRT; 430 AA.
AC Q89H48;
DT 01-JUN-2003 (TremBLrel. 24, Created)
DT 01-JUN-2003 (TremBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TremBLrel. 25, Last annotation update)
DE BL16147 protein.
GN BL16147.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA 110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneko T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasamoto S., Watanabe A., Idesawa K., Iriguchi M., Kawashima K.,
RA Kohara M., Matsumoto M., Shimpo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110."
RL DNA Res. 9:189-197 (2002).
DR EMBL; AP005957; BAC51412.1; -.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase M20.
DR InterPro; IPR001261; Peptidase M20_A.
DR Pfam; PF01546; Peptidase M20; 1.
DR PROSITE; PS00758; ARGE_DAPE_CPG2_1; 1.
KW Complete proteome.
SQ SEQUENCE 430 AA; 45954 MW; E3BD103B9599E532 CRC64;

Query Match 34.7%; Score 710.5; DB 16; Length 430;
Best Local Similarity 38.5%; Pred. No. 1.3e-36;
Matches 154; Conservative 73; Mismatches 168; Indels 5; Gaps 3;

Qy 11 IEKEIRELSRPSAEGPGVTRLTYPTEHAARRETLIAAKAAALSVRDALGNIIIGREGT 70
Db 22 IVRRINVLGAISEETDKLTRYLSKELRAAADLILGMWREAGMSAHLDAIGNVCGRYGE 81
Qy 71 DPFLPAIAGVSHFDSVRNGMFDGTAGVVCALAAARVLENGYVNRHPF--EFIAIVERE 128
Db 82 RFGAPCLMLGSHYDTRDAGKWDGPGLVITAIACVADLNRRG--KELFAIEVIGFADEE 139
Qy 129 GARFSSGMLGGRAIAGLVADRELDSLVDEGVSVRQAATAFGLKPGELQAAARSADLRA 188
Db 140 GVRFASTLLGSRVAVGTDFESVLNTR--DRDGVSMRDALVKFGLDHPDHGAAARARRELLA 198
Qy 199 FIELHIEQGPILQEQIEIGVTVTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPAALM 248
Db 199 YLELHIEQGPVLEAQNLPVGVVTAIATAGATRLAVRLNGMAGHAGTVPMLRRDALTGAAEC 258
Qy 249 VRENVRFNEIADGTATVGHLTVPAGGQVQVGEVDFTLRLSPHEESLRVLIDRISVM 308
Db 259 IGATIQFCRTDEGLVGVTVGIQIARPGATNIPGEVSTIDNRAFTDMHRKRAVADVVRQ 318
Qy 309 VGEVASQAGVAADVDFEFLNLSVQLAPTMTVDVAVREAAASALQFTHRDISSGAGHDSMFIAQ 368
Db 319 IEAIAKRQALQLDVTHTENRTAPCASWLKDQIAQIAAAGSVSVFDPSPGAGHDMGAMID 378
Qy 369 VTDVGMVFPVPSRAGRSHVPEWTFDDLRKGTGVVLRMK 408
Db 379 IADVGMIFVRCRGGVSHHPDEHVELADADAGARVLLRVIE 418
```

```
Query Match 41.6%; Score 853; DB 16; Length 414;
Best Local Similarity 43.2%; Pred. No. 1.5e-45;
Matches 173; Conservative 73; Mismatches 152; Indels 2; Gaps 2;

Qy 10 RIEKEIRELSRFA-EGGVTRLTYTPPEHAARETLIAAKAAALSVDALGNIIGRR 68
Db 7 RIKKHIEINLDRFTATPGGTTTLTYSKEDLGARNYLKEEMAKVGLTVSEDAIGNIYGRLE 66
Qy 69 GTPPELPAIAGVSHFSDVRNGGFDGTAGVVCALAEARVMLENGYVNRHHPFEIATVEEE 128
Db 67 GDNPDIPAVIAGVSHFSDVRNGGFDGPAGVITGLVASVPHEOQIKPHFPLEIATVMEVEE 126
Qy 129 GARFSSGMLGGRATAGLVADRELDLSLVDGVSVRQAATAFGLKPGKGLQAARSADLRA 188
Db 127 GSRFGAGLLASRTITGKVTYKEMLEHMDKINGVTAEEAMNLGPDANGVHTAIRSKSVKA 186
Qy 189 FIELHIEQGPILQEIQEIGVWTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPAALM 248
Db 187 FIELHIEQGPVLENANEDVALDVTVGLTEIKVTVKQAGHAGTTPMLERKDALSAVQI 246
Qy 249 VREVNRFVNEIADGTATVGHLTVPAGGQNVGPEVDFTDLRSPHEESLRVLIDRISVM 308
Db 247 LNLKPELAIQEGGTVLTIGLNVYPNGANVPKNKVFTVDIRAKDEIHVQNTLEKTKV 306
Qy 309 VGVASQAGVAADVDEFFNLSPVLQAPTMVDVAREASALQFTHRDISSGAGHDSMFIAQ 368
Db 307 I-QAFENKNGIMCEDMLYKPTHLSHTEIHOALTESADKUGLKRYRTMVSGAGHDAMIFAS 365
Qy 369 VTDGVMFVPSRAGRSHVPEWTFDRLRGTEVLRVMK 408
Db 366 LTEVGLIFVPSHGKISHAPEWTDYDKLQKIEVLETVK 405

RESULT 4
Q8EKY8 PRELIMINARY; PRT; 413 AA.
AC Q8EKY8;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-carbamyl-L-amino acid amidohydrolase (EC 3.5.1.-).
GN OB3444.
OS Oceanobacillus iheyensis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.
OX NCBI_TaxID=182710;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HTE831 / DSM 14371 / JCM 11309;
RX MEDLINE=2220767; PubMed=12235376;
RA Takami H., Takaki Y., Uchiyama I.;
RT "Genome sequence of Oceanobacillus iheyensis isolated from the Iheya
RT Ridge and its unexpected adaptive capabilities to extreme
RT environments.";
RL Nucleic Acids Res. 30:3927-3935(2002).
DR EMBL; AP004604; BAC15400.1; -.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase M20.
DR Pfam; PF01546; Peptidase M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 413 AA; 44900 MW; 9573E6CF75EC93D CRC64;

Query Match 41.3%; Score 847; DB 16; Length 413;
Best Local Similarity 41.7%; Pred. No. 3.5e-45;
Matches 168; Conservative 84; Mismatches 147; Indels 4; Gaps 3;

Qy 9 ARIEKEIRELSRFA-EGGVTRLTYTPPEHAARETLIAAKAAALSVDALGNIIGRR 67
Db 6 ARIEHIDAISKFTATPGGTTTLTYSKQDKQAKYIKEMAEYDLEVEDSGFNIFGKL 65
Qy 68 EGTDPPELPAIAGVSHFSDVRNGGFDGTAGVVCALAEARVMLENGYVNRHHPFEIATVEE 127
```

```
Db 66 EGTIKDAPSLTIGSHFSDVRNGGSDGPGVAGVLEVAALFNQNNLKPYPLEVALIEE 125
Qy 128 EGARFSSGMLGGRATAGLVADRELDLSLVDGVSVRQAATAFGLKPGELQAARSADLR 187
Db 126 EGSRFGGGLMGSGMTGLREEDFNKLTDTYNGITTEAMREIGLDPS--LPKRDSETIK 183
Qy 188 APTIELHIEQGPILQEIQEIGVWTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPAAL 247
Db 184 SYLELHIEQGPILDEKKIPGVVETIVGLTQPEITVEGQAGHAGTTPMDHRSDALVTAQ 243
Qy 248 MVREVNRFVNEIADGTATVGHLTVPAGGQNVGPEVDFTDLRSPHEESLRVLIDRISV 307
Db 244 MIAQIPSLAVEEGEGTVATTGKLVNFPNGSNVPSKTVFTVDIRSGKEBHIQHVDKLHE 303
Qy 308 MVGEVASQAGVAADVDEFFNLSPVLQAPTMVDVAREASALQFTHRDISSGAGHDSMFIA 367
Db 304 MANSY-NRDKITITISQQLMEPKAMNPDIALLKETSSFDIPYCSMNSGAGHDAMVLA 362
Qy 368 QVTDGVMFVPSRAGRSHVPEWTFDRLRGTEVLRVMKAL 410
Db 363 EVDVGMLFIPSKDGVSHVPEWSDSLDIKAVEIDFAAAKKL 405

RESULT 5
Q9KET8 PRELIMINARY; PRT; 414 AA.
AC Q9KET8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE N-carbamyl-L-amino acid amidohydrolase.
GN BH0761.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RL Nucleic Acids Res. 28:4317-4331(2000).
DR EMBL; AP001509; BAB04480.1; -.
DR PIR; A83745; A83745.
DR GO; GO:0016787; F:hydrolase activity; IEA.
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase M20.
DR Pfam; PF01546; Peptidase M20; 1.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 414 AA; 45115 MW; 59D9401D81328A11 CRC64;

Query Match 37.2%; Score 762.5; DB 16; Length 414;
Best Local Similarity 39.6%; Pred. No. 6.9e-40;
Matches 161; Conservative 82; Mismatches 153; Indels 11; Gaps 4;

Qy 10 RIEKEIRELSRFSABGP---GVTRLTYTPPEHAARETLIAAKAAALSVDALGNIIG 65
Db 11 LRWRTIMELAEIGGTSPIDVHGVTRLSLTKELRARQYVIDLMKOSGLSVQDVGNIIG 70
Qy 66 REGTDPPELPAIAGVSHFSDVRNGGFDGTAGVVCALAEARVMLENGYVNRHHPFEIATV 125
Db 71 KLEGTDLLELPVMTGSHIDSVPHGGRFGDTGLVGLGAIEAVRTMKEAGIKLKHISIEIVFT 130
Qy 126 EREGARFSSGMLGGRATAGLVADRELDLSLVDGVSVRQAATAFGLKPGELQAARSAD 185
Db 131 DEEGARFGAGTGGKMGAGELTETFTT-SLADDKGYTYREAFLEAANLNPITYKQAIRSDEQ 189
Qy 186 LRAFTIELHIEQGPILQEIQEIGVWTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPA 245
```

Matches 411; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTLQAAQAARIKEIRELSRFSAGPGVTRLTYTPPEHAARETLTAAMKAALSVREDAL 60
 DB 1 MTLQAAQAARIKEIRELSRFSAGPGVTRLTYTPPEHAARETLTAAMKAALSVREDAL 60

QY 61 GNIIGRECTDPELPAIAGVSHFDSVRNGMGFDGTAGVVCALAEAAARVMLENGYVNRHPPE 120
 DB 61 GNIIGRECTDPELPAIAGVSHFDSVRNGMGFDGTAGVVCALAEAAARVMLENGYVNRHPPE 120

QY 121 FTAIVVEEGARSSGMLGRATAGLVADRELSDLVDEDDGVSVRQAATAFGLKPEGLQAAA 180
 DB 121 FTAIVVEEGARSSGMLGRATAGLVADRELSDLVDEDDGVSVRQAATAFGLKPEGLQAAA 180

QY 181 RSAADLRAPFIELHIEOGPILEQEQIEIGVTSIVGVRALRVAVKGRSDHAGTTPMHLROD 240
 DB 181 RSAADLRAPFIELHIEOGPILEQEQIEIGVTSIVGVRALRVAVKGRSDHAGTTPMHLROD 240

QY 241 ALVPAALMVREVRNFVNEIADGTATVGHLTVAPEGGNQVPGVEVFTLDRSPHEESLRV 300
 DB 241 ALVPAALMVREVRNFVNEIADGTATVGHLTVAPEGGNQVPGVEVFTLDRSPHEESLRV 300

QY 301 LIDRISVMVGEASQAGVAADVDEFPNLSVQLAPTMVDVAREASALQFTHRDSSGAG 360
 DB 301 LIDRISVMVGEASQAGVAADVDEFPNLSVQLAPTMVDVAREASALQFTHRDSSGAG 360

QY 361 HDSMFIAQVTDVGMVFPVRAGRSHVPEEWTDFDLRKGTGVVLRVMKALDR 412
 DB 361 HDSMFIAQVTDVGMVFPVRAGRSHVPEEWTDFDLRKGTGVVLRVMKALDR 412

RESULT 2

Q8Y9J1 PRELIMINARY; PRT; 423 AA.

AC Q8Y9J1
 DT 01-MAR-2002 (TremBLrel. 20, Created)
 DT 01-MAR-2002 (TremBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Hypothetical protein lmo0537.
 GN LMO0537.
 OS Listeria monocytogenes.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1639;
 RN [1]
 RC STRAIN=EGD-e / Serovar 1/2a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glauber P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species.";
 RL Science 294:849-852(2001).
 DR EMBL; AL591975; CAC98616.1; -.
 DR PIR; AB1142; AB1142.
 DR ListList; LMO0537; -.
 DR GO; GO:0005622; C:intracellular; IEA.
 DR GO; GO:0008237; F:metallopeptidase activity; IEA.
 DR GO; GO:0003700; F:transcription factor activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR000524; HTH GntR.
 DR InterPro; IPR002933; Peptidase M20.
 DR Pfam; PF01546; Peptidase M20; I.
 DR PROSITE; PS00043; HTH_GNTR_FAMILY; 1.
 KW Hypothetical protein; Complete proteome.

QY 10 RIEKIRELSRFSAGPGVTRLTYTPPEHAARETLTAAMKAALSVREDALNIGRRE 68
 DB 7 RIKKHIEKLDFTATPGQGTTRLTVSKEDLDARNYLKQEMAKVGLTVSEDAIGNIYGRLE 66

QY 69 GTDPELPAIAGVSHFDSVRNGMGFDGTAGVVCALAEAAARVMLENGYVNRHPPEFIAVSEE 128
 DB 67 GESPDPPIAVIAGVSHFDSVRNGMGFDGTAGVVCALAEAAARVMLENGYVNRHPPEFIAVSEE 126

QY 129 GARFSSGMLGRATAGLVADRELSDLVDEDDGVSVRQAATAFGLKPEGLQAAAADLR 188
 DB 127 GSRFAGALLASRAITGKVTTEMLHEMKDIDGITATDAMAKLGFNANRVHRAIRTKESVKA 186

QY 189 FIELHIEOGPILEQEQIEIGVTSIVGVRALRVAVKGRSDHAGTTPMHLRQDALVPAALM 248
 DB 187 FIELHIEOGPILENASSEDVALVDTVVGLTQIKVTVGQAGHAGTTPMLDRKDALVSAVEI 246

QY 249 VREVRNFVNEIADGTATVGHLTVAPEGGNQVPGVEVFTLDRSPHEESLRVLRISVM 308
 DB 247 LGQPELAIQGGGFTLVGKLVVYFNGANVDPKVIPTVDIRAKDEIHVQNTLAKTKEI 306

QY 309 VGEASQAGVAADVDEFPNLSVQLAPTMVDVAREASALQFTHRDSSGAGHDSMFIAQ 368
 DB 307 I-QATEKNGITCEIKDMYIQQTHLSKEIHALTESADQLGFKYRTMTVSGAGHDMIFAS 365

QY 369 VTDGMVFPVRAGRSHVPEEWTDFDLRKGTGVVLRVMK 408
 DB 366 LTEVGLFVPSHNGISHAPEEWTDYDKLGKIEVVLTKVK 405

RESULT 3

Q92EB9 PRELIMINARY; PRT; 414 AA.

AC Q92EB9
 DT 01-DEC-2001 (TremBLrel. 19, Created)
 DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TremBLrel. 24, Last annotation update)
 DE Hypothetical protein lin0541.
 GN LIN0541.
 OS Listeria innocua.
 OC Bacteria; Firmicutes; Bacillales; Listeriaceae; Listeria.
 OX NCBI_TaxID=1642;
 RN [1]
 RC STRAIN=CLIP 11262 / Serovar 6a;
 RC MEDLINE=21537279; PubMed=11679669;
 RA Glauber P., Frangeul L., Buchrieser C., Rusniok C., Amend A.,
 RA Baquero F., Berche P., Bloeker H., Brandt P., Chakraborty T.,
 RA Charbit A., Chetouani F., Couve E., de Daruvar A., Dehoux P.,
 RA Domann E., Dominguez-Bernal G., Duchaud E., Durant L., Dussurget O.,
 RA Entian K.-D., Fsihi H., Garcia-del Portillo F., Garrido P.,
 RA Gautier L., Goebel W., Gomez-Lopez N., Hain T., Hauf J., Jackson D.,
 RA Jones L.-M., Kaerst U., Krest J., Kuhn M., Kunst F., Kurapkat G.,
 RA Madueno E., Maitournam A., Mata Vicente J., Ng E., Nedjari H.,
 RA Nordsiek G., Novella S., de Pablo B., Perez-Diaz J.-C., Purcell R.,
 RA Remmel B., Rose M., Schlueter T., Simoes N., Tierrez A.,
 RA Vazquez-Boland J.-A., Voss H., Wehland J., Cossart P.;
 RT "Comparative genomics of Listeria species.";
 RL Science 294:849-852(2001).
 DR EMBL; AL596165; CAC95773.1; -.
 DR PIR; AB1500; AB1500.
 DR ListList; LIN00541; -.
 DR GO; GO:0008237; F:metallopeptidase activity; IEA.
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
 DR InterPro; IPR002933; Peptidase M20.
 DR Pfam; PF01546; Peptidase M20; I.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 414 AA; 45033 MW; A649E28B344A2A25 CRC64;

GenCore version 5.1.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 18:17:50 ; Search time 45 Seconds
(without alignments)
2888.744 Million cell updates/sec

Title: US-10-045-063-2

Perfect score: 2049

Sequence: 1 MTLQKAQAARIEKRELSR.....FDDLKRGTEVLRVVKALDR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 25:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phase:*
- 10: sp_plant:*
- 11: sp_rodent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriap:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2044	99.8	412	2 Q9F464	Q9f464 arthrobacte
2	858	41.9	423	16 Q8Y9J1	Q8y9j1 listeria mo
3	853	41.6	414	16 Q92EB9	Q92eb9 listeria in
4	847	41.3	413	16 Q8EKY8	Q8eky8 oceanobacil
5	762.5	37.2	414	16 Q9KET8	Q9ket8 bacillus ha
6	720.5	35.2	409	2 Q8GGG5	Q8ggg5 geobacillus
7	710.5	34.7	430	16 Q89H48	Q89h48 bradyrhizob
8	691	33.7	420	16 Q7WJR1	Q7wjrl bordetella
9	687	33.5	420	16 Q7WAL0	Q7wal0 bordetella
10	663	32.4	412	16 Q9CP93	Q9cp93 pasteurella
11	645.5	31.5	423	16 Q8PDQ1	Q8pdq1 xanthomonas
12	642.5	31.4	421	16 Q98LM4	Q98lm4 rhizobium 1
13	642.5	31.4	430	16 Q8ZBY1	Q8zby1 yersinia pe
14	642.5	31.4	431	16 Q8DI76	Q8di76 yersinia pe
15	639.5	31.2	427	16 Q91673	Q91673 pseudomonas
16	636	31.0	441	10 Q9FIY0	Q9fiy0 arabidopsis

17	636	31.0	476	10 Q8VXY9	Q8vxy9 arabidopsis
18	631.5	30.8	427	16 Q88FQ3	Q88fq3 pseudomonas
19	629	30.7	454	16 Q7UHC4	Q7uhc4 rhodopirell
20	625.5	30.5	425	16 Q88Q81	Q88q81 pseudomonas
21	617.5	30.1	423	16 Q8PQM2	Q8pqm2 xanthomonas
22	612.5	29.9	427	16 Q9A561	Q9a561 caulobacter
23	606	29.6	407	16 Q82ZQ2	Q82zq2 enterococcu
24	605.5	29.6	438	16 Q7U3I0	Q7u3i0 synechococc
25	604	29.5	423	16 Q7WLK0	Q7wlk0 bordetella
26	603.5	29.5	420	16 Q7W7T2	Q7w7t2 bordetella
27	601.5	29.4	420	16 Q7WL72	Q7wl72 bordetella
28	600	29.3	423	16 Q7WPU1	Q7wpu1 bordetella
29	599	29.2	414	16 Q985I5	Q985i5 rhizobium 1
30	596.5	29.1	420	16 Q7WVF6	Q7wvf6 bordetella
31	591.5	28.9	410	16 Q8ELE3	Q8ele3 oceanobacil
32	583.5	28.5	416	16 Q98K54	Q98k54 rhizobium 1
33	575	28.1	420	2 Q8VUL6	Q8vul6 pseudomonas
34	563.5	27.5	416	16 Q89R53	Q89r53 bradyrhizob
35	562	27.4	415	16 Q8UCU8	Q8ucuc8 agrobacteri
36	562	27.4	416	16 Q9RV75	Q9rv75 deinococcus
37	558.5	27.3	410	16 Q984M3	Q984m3 rhizobium 1
38	554.5	27.1	414	16 Q92U74	Q92u74 rhizobium m
39	548	26.7	416	16 Q92MZ4	Q92mz4 rhizobium m
40	547.5	26.7	406	16 Q893C9	Q893c9 clostridium
41	547.5	26.7	415	16 Q7WJ52	Q7wj52 bordetella
42	547.5	26.7	415	16 Q7WA23	Q7wa23 bordetella
43	546.5	26.7	415	16 Q8G2N9	Q8g2n9 brucella su
44	545.5	26.6	415	16 Q7VXC4	Q7vxc4 bordetella
45	536.5	26.2	415	16 Q8YF79	Q8yf79 brucella me

ALIGNMENTS

RESULT 1

Q9F464.
ID Q9F464 PRELIMINARY; PRT; 412 AA.
AC Q9F464;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE L-N-carbamoylase Hyc.
GN HYUC.
OS Arthrobacter aureus.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcineae; Micrococcaceae; Arthrobacter.
OX NCBI_TaxID=43663;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM3747;
RX MEDLINE=99210756; PubMed=10194852;
RA Wilms B., Wiese A., Syltatk C., Mattes R., Altenbuchner J.,
RA Pietzsch M.;
RA "Cloning, nucleotide sequence and expression of a new L-N-carbamoylase
RT gene from Arthrobacter aureusens DSM 3747 in E. coli.";
RL J. Biotechnol. 68:101-113(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=DSM3747;
RX MEDLINE=20403588; PubMed=10949312;
RA Wiese A., Pietzsch M., Syltatk C., Mattes R., Altenbuchner J.;
RT "Hydantoin racemase from Arthrobacter aureusens DSM 3747: heterologous
RT expression, purification and characterization.";
RL J. Biotechnol. 80:217-230(2000).
DR EMBL; AF146701; AAC02131.1; -
DR GO; GO:0008237; F:metallopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR002933; Peptidase M20.
DR Pfam; PF01546; Peptidase_M20; I
SQ SEQUENCE 412 AA; 44084 MW; 6E8B76F923FBA2CA CRC64;
Query Match 99.8%; Score 2044; DB 2; Length 412;
Best Local Similarity 99.8%; Pred. No. 3e-120;

Search completed: May 3, 2004, 18:21:00
Job time : 21 secs

Db 69 LGGTGLGHTRWATGRPTDRNAHPRDRAAGKIAVHNGIIENFAPLRAELEAGV---- 124
 Qy 112 GYVNRHPFEFTIAVEEGA-----RFSGMLGGRAIAGLVADRE-----LDSLVE 157
 Db 125 -----EFASDTEVAHVILVARQYTOGDTAGDPPASVLAVLQRLGCHFTLVASADD 176
 Qy 158 DOVSU---RQAATAGLKGPELOAARSADURAFIELHIEGPILEQEQIBIGVWTSIV 214
 Db 177 PGTIVAARRSTPLVLGIGDGENFV---GSDVAAFIE-HTRDAVELGDDQ---AVVLTD 228
 Qy 215 GVRA-----LRVAVKGRSDHAGTTPMHLRQDALVPAALM--- 248
 Db 229 GYRITDPAGNDHLEAGRDPRFPHIDWDLNAEKGVDY-----FMLKEIAEOPSAVDTL 283
 Qy 249 -----VRENRVFNVEIADGTVATVGHILT-VAPGGGNQVPGVEOPT 287
 Db 284 LGHFDKNRVLDEQRLSDQELREIDK-VFIVACGTAYHSGLLAKYAIEHWTLPVEVELA 342
 Qy 288 LDL--RSPHEESLRVLDRISVMVGEVASOAGVAADVDEFFNLSPVOLAPTIV---D 339
 Db 343 SEFRYRDP-----VLDRLSTLVI--AISQGETADTLEAVRHAKTKAKVLAICNTNGS 393
 Qy 340 AVREAASALQFTHRDISSGAGHDSMFIAQVTD--VGMVFPVPSRAGSHVPEEWTDFDDL 396
 Db 394 QIPREDAVLYTRAGPEIGVAAKTFLAQIAANYLVGLALAQAR-GTKYPDEVAEYREL 452
 Qy 397 RKGTEVLRVMKALD 411
 Db 453 EAMPDLIKREVLAGMD 467

RESULT 14

TRAA RHISN STANDARD; PRT; 1102 AA.
 AC P55478.
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Probable conjugal transfer protein traa.
 GN TRAA OR Y4DS.
 OS Rhizobium sp. (strain NGR234).
 OG Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 OX NCBI_TaxID:394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97305956; PubMed=9163424;
 RA Freiberg C.A., Fellay R., Bairoch A., Broughton W.J., Rosenthal A.,
 RA Perret X.;
 RT "Molecular basis of symbiosis between Rhizobium and legumes.";
 RL Nature 387:394-401(1997).
 CC -!- SIMILARITY: Belongs to the mobA/mobL family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; AE000069; AAB91648.1; -;
 DR InterPro; IPR005053; MobA MobL.
 DR Pfam; PF03389; MobA MobL; 1.
 KW Conjugation; ATP-binding; Plasmid.
 FT NP_BIND 404 411 ATP (POTENTIAL).
 SQ SEQUENCE 1102 AA; 123033 MW; EC17A5528F7B9CAF CRC64;

Query Match

Best Local Similarity 5.6%; Score 114; DB 1; Length 1102;

Matches 106; Conservative 48; Mismatches 150; Indels 136; Gaps 23;

Qy 3 LOKQAARIEKE-----IRELSRSAB-GPGVTRLTYYTPEHAARETLIA 46
 Db 315 LQSPKTLERERIAFATGIRAPAKYTTRELIRLEAENGSRAIWLSSRSSHGVRKEVLEA 374
 Qy 47 AM-KAALS-VREDALGNIIGRRECTDPELPAIAGVSHFDSVRNGMGDFGTAGVVVCALEA 104
 Db 375 AFSRHSRLSDQKTAIEHVAQ-----AERTAAVIG-----RAG-----AGKTTMKA 416
 Qy 105 ARVMLENGYVNRHPFEFTIAVEEGARFSSGMLGGRAIAGL-----VADRELDG----- 153
 Db 417 AR-----EAMEAAGYRVVGGALAGKAAEGLEKEAGIASRTLSSELMRW 460
 Qy 154 -----LVDEDC-VSVRQ-----AATAFG-----LKPGELOAARSA 183
 Db 461 EGRKQLDDKTFVLDEAGMWSRQWALEFETATKAGAKLVLVGDPEQLOPTEAGAAFRAI 520
 Qy 184 ADLRAFIELHI-----EOGPILEQEQIBIGVVTISVGV-----RALRVAVKGRS----- 227
 Db 521 ADRTGYAELETTYRQEQWMCASLDLARGNVKVVDTYRANGRMRSSELKAEAVQNLIA 580
 Qy 228 -----DHAGTTPM--HLRQDA-LVPAALMVRVNRVFNVEIADGTATVGHILTVPAGGN 278
 Db 581 DWDRDYDPTKTTLLHLRRDVRMLNQWAKLVERGIVDAGFSFKAEDGNRRFAP--GD 638
 Qy 279 QV-----PGEVDFTLDRSPHEESLRVLDRISVMVGEVASOAGVAADVDEFFNLSPVOL 333
 Db 639 QIVFLKNEGALGVKNGMRGKVVEAAQ---NRIVAEIGEVEHRRQVMVE-SRFYNNLDHGY 694
 Qy 334 APTM-----VDAVREAS 346
 Db 695 ATTIKSGAIVDRVKVLAS 714

RESULT 15

ERV3 SACER
 ID ERV3 SACER STANDARD; PRT; 3172 AA.
 AC Q03133; Q54097; Q99270;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DE Erythronolide synthase, modules 5 and 6 (EC 2.3.1.94) (ORF 3) (6-
 DE deoxyerythronolide B synthase III) (DEBS 3).
 DE ERYA.
 OS Saccharopolyspora erythraea (Streptomyces erythraeus).
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Pseudonocardiaceae; Pseudonocardiaceae; Saccharopolyspora.
 OX NCBI_TaxID=1836;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 2338;
 RX MEDLINE=91043075; PubMed=2234082;
 RA Cortes J., Haydock S.F., Roberts G.A., Bevitt D.J., Leadlay P.F.;
 RT "An unusually large multifunctional polypeptide in the erythromycin-
 RT producing polyketide synthase of Saccharopolyspora erythraea.";
 RL Nature 348:176-178(1990).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91220065; PubMed=2024119;
 RA Donadio S., Staver M.J., McAlpine J.B., Swanson S.J., Katz L.;
 RT "Modular organization of genes required for complex polyketide
 RT biosynthesis.";
 RL Science 252:675-679(1991).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRRL 2338;
 RX MEDLINE=92155230; PubMed=1740151;
 RA Bevitt D.J., Cortes J., Haydock S.F., Leadlay P.F.;
 RT "6-deoxyerythronolide-B synthase 2 from Saccharopolyspora erythraea.
 RT Cloning of the structural gene, sequence analysis and inferred domain
 RT structure of the multifunctional enzyme.";
 RL Eur. J. Biochem. 204:39-49(1992).
 CC -!- CATALYTIC ACTIVITY: 6 malonyl-CoA + propionyl-CoA = 7 CoA + 6-
 CC deoxyerythronolide B.

QY 131 -----RFSSGMLGGRAIGLVAD-RELD---SLV-----DEGVSV---ROAATAFGLKPG 174
 Db 137 HLVAQYACAGTAGDFGVGSLAVLRRLQGHFTLVFANADEPGTIVAARRSTPLVLIGDG 196
 QY 175 ELQAAARSADLRARTELHIEOGPILEOEOIIGVVTISGVRA-----218
 Db 197 EMFV-----GSDVAAFIE-HTRQAVELGQDQ---AVVITADGRIISDFDGNDDAVNARTFH 248
 QY 219 ----LRVAVKGRSDHAGTTPMHLRQDALVPAAL-----MVREVRNFV 256
 Db 249 IDMDLAAAEKGGVEYFMLEIAEQADPAVVDTLHGFTGRIVLDSQRLSDQELREIDK-V 307
 QY 257 NEIAGTV-----ATVGHTVAPGGNGQVGPGEVDFTLDL--RSPHEESLRVLIDRIS 306
 Db 308 FVAVCGCTAYHSGLLAKYTTIEHT-----RLPVEVELASEFRYRDP-----VLDIRST 353
 QY 307 VMVGEVASQAGVAADVDEFFNLSPVQLAPTMV-----DAVREASALQFTHRDSSGAG 360
 Db 354 LVV--AISQSGETADTLEAVRHAKOKAKVLAICNTNGSQIPRECDAVLYTRAGPEIGVA 411
 QY 361 HDSMFIAQVTD---VGMVFPVPSRAGRSHVPEEWTDFDRLKGTGVVLRVMKAL 410
 Db 412 STKTFLAQVAANYLLGLALAQAR-GTKYPDEVQREYRELEAMPDLVARVIAGM 463

RESULT 11
 VGLH HSVBC STANDARD; PRT; 842 AA.
 AC P2759;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 01-AUG-1992 (Rel. 23, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glycoprotein H precursor.
 GN OR UL22.
 OS Bovine herpesvirus type 1 (strain Cooper).
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Alphaherpesvirinae; Varicellovirus.
 OX NCBI_TaxID=10323;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Meyer A.L., Petrovskis E.A., Duffus W.P.N., Thomsen D.R.,
 RA Post L.E., Apr-1991 to the EMBL/GenBank/DBJ databases.
 RL Submitted (APR-1991) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: GH AND GL FORM A COMPLEX THAT INDUCES NEUTRALIZING AND
 CC SYNCTYTA INHIBITING ANTIBODIES (BY SIMILARITY).
 CC -!- SIMILARITY: Belongs to the herpesviruses glycoprotein H family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X58867; CA441677.1; -
 DR EMBL; Z78205; CAB01604.1; -
 DR EMBL; AJ004801; CA006112.1; -
 DR PIR; S18462; S18462.
 KW Glycoprotein; Transmembrané; Signal.
 FT SIGNAL 1 25 POTENTIAL.
 FT CHAIN 26 842 GLYCOPROTEIN H.
 FT CARBOHYD 77 77 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 112 112 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 617 617 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 666 666 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 760 760 N-LINKED (GLCNAC. .) (POTENTIAL).
 FT CARBOHYD 783 783 N-LINKED (GLCNAC. .) (POTENTIAL).
 SQ SEQUENCE 842 AA; 88375 MW; BA90759A74715F98 CRC64;

Query Match 5.7%; Score 117.5; DB 1; Length 842;
 Best Local Similarity 21.7%; Pred. No. 2.9; Mismatches 171; Indels 143; Gaps 19;
 Matches 100; Conservative 47;

QY 9 ARIEKEIRLSRFS-----AEGP-----GVTRL-----TYTPEH 37
 Db 270 APADANVRMITAFNGGGGAFPPPPGPAAGPQRRAYVIGYNSRLDSHMYLTMREVASYANEP 329
 QY 38 AAARETLTAAKKAALSVRDALGNIIIGRRGTDELP-----AIAVSHFSDSVRNGGM 91
 Db 330 ADFAHLTAHREAFMLUREAAAR-RGPSAGPAPNAAHYAYRVAARLGLALSALTEGAL 388
 QY 92 FDGTAGVVCAL-----AARVMLENGVNRHPPEFIAIVEEGARFSSGMLGGRA 141
 Db 389 AD---GYVLAELVDLDYHLKLSRVLLGAGL-----GCAANGVRVART 429
 QY 142 TAGLVADREL--DSLVDDEGVSVRQAATAFGLKPELOAAARSAAADLRAFTELHIEQGP 199
 Db 430 IAQLAVPRELPDAFIPPEPAGAALSVVARGKLRVAVAFSGPDAPLAARLAH-----483
 QY 200 LEQEQIEIGVVTISGV-----RALRVAVKGRSDHAGTTPMHLRQDALVPA 246
 Db 484 -----GVVSDLYDAFLRGELTWGPPMRHALFFAFAASAPPAQAQALELARDVTRKCT 535
 QY 247 LMVREVRNFVNEIADGTVATVGHLTVA-----GGGNQVPEVDF-TLDLRSPE 295
 Db 536 AM-----CTAGHATAAALDLEEVYAHVGGG--AGDAGFELLDAFSCM 577
 QY 296 ESLRV-LIDRISVMVGEVASQAGVAADVDEFFNLSPVQLAPTMVDVAREASALQFTHRD 354
 Db 578 ASFRLLDLEEAHV--DVLSAVPARAALDAWLEAQAPAAAAPNL-----SAAALGMLGRG 629
 QY 355 ISSGAGHDSMTIAQV-----TDVGMVFPVPSRAGRSHV 386
 Db 630 GLFGPAHAAALAPELFAAPCGGWGAGAAVAIVPVAPNASYV 670

RESULT 12
 CARB_HELPJ STANDARD; PRT; 1085 AA.
 AC Q9ZKT2;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Carbamoyl-phosphate synthase large chain (EC 6.3.5.5) (Carbamoyl-
 DE phosphate synthetase ammonia chain).
 GN CARB OR JHP0853.
 OS Helicobacter pylori J99 (Campylobacter pylori J99).
 OC Bacteria; Proteobacteria; Epsilonproteobacteria; Campylobacteriales;
 OC Helicobacteraceae; Helicobacter.
 OC NCBI_TaxID=855963;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99120557; PubMed=9923682;
 RA Alm R.A., Ling L.-S.L., Moir D.T., King B.L., Brown E.D., Doig P.C.,
 RA Smith D.R., Noonan B., Guild B.C., deJonge B.L., Carmel G.,
 RA Tummino P.J., Caruso A., Uria-Nickelsen M., Mills D.M., Ives C.,
 RA Gibson R., Merberg D., Mills S.D., Jiang Q., Taylor D.E., Vovis G.F.,
 RA Trust T.J.;
 RT "Genomic sequence comparison of two unrelated isolates of the human
 RT gastric pathogen Helicobacter pylori.";
 RL Nature 397:176-180(1999).
 CC -!- CATALYTIC ACTIVITY: 2 ATP + L-glutamine + CO(2) + H(2)O = 2 ADP +
 CC phosphate + L-glutamate + carbamoyl phosphate.
 CC -!- COFACTOR: Binds 3 manganese ions per subunit (By similarity).
 CC -!- PATHWAY: Arginine biosynthesis.
 CC -!- SUBUNIT: Composed of two chains; the small (or glutamine) chain
 CC promotes the hydrolysis of glutamine to ammonia, which is used by
 CC the large (or ammonia) chain to synthesize carbamoyl phosphate (By
 CC similarity).
 CC -!- SIMILARITY: Belongs to the carB family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its

RL Proc. Natl. Acad. Sci. U.S.A. 100:7877-7882(2003).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z77137; CAB00904.1; -
DR EMBL; AE007006; AAK45576.1; -
DR EMBL; BX248338; CAD94170.1; -
DR PIR; F70755; F70755.
DR TIGR; MT1315; -
DR Tuberculist; rv1278; -
DR Hypothetical protein; Complete proteome.
KW SEQUENCE 875 AA; 93350 MW; 366C580A3AAC97A2 CRC64;
SQ
Query Match 5.9%; Score 121; DB 1; Length 875;
Best Local Similarity 24.2%; Pred. No. 1.8;
Matches 109; Conservative 40; Mismatches 155; Indels 146; Gaps 22;
QY 26 PGVRLTYTPPEHAARETL-----IAAMKAAALSVRDGLGNIIGREGTDPQL 74
DB 465 PGATALDFQAKYAAQOELADALAGGEVADLAARSADLCRE-----LLSRDQLTATL 519
QY 75 PATAVSHFVSUNG-----GMFDGTA-GVVCALAEARVLMENGVNRRHP 118
DB 520 AGLCGDEQVQLSRLEQOLCAGOPAEILDVSTDTATARAELDAVEAARIAAEKDCETRRQ 579
QY 119 F-----EFIAIVEEGARFSGMLGG-----RATAG---LVADRELD--- 152
DB 580 IAAGAARRLAETTRATVQLNQAASAEALGAAMTRILACRASVGDDLEAKAEADLRVL 639
QY 153 -----SLVSDGVSVRQAATAFGLKPGELQAAARSAAADLRAFTLHTEQGPILQEQQ 204
DB 640 QTAEQRVIDLADE-----LAATAPDAVAELAEADAVELLR---ERHDEATRALH--- 687
QY 205 IEIGVTVTSIVGVRAALRVAVKGRSDHAGT-----TPMHLRQDALVPALMVRVN----- 253
DB 688 -EVGVLSVFTGQR-----KGKLDAAETEREHAASHHARVGRRAARALLRSVWARHRDT 742
QY 254 ---RFVN---ETADGTVATVGHITVAPGGNQVPGVEVDTLRLSPHEESLVLDL--- 304
DB 743 TRURYVEPYRAEL-----HRLGRVPFGPSFEVEVDTLRLS-----RTLDDRTV 787
QY 305 -ISVMVGEVASQAGVAADVDEFFNLSPVQLAPTWDAVREASALQFTHRDSSGAGHDS 363
DB 788 PYECLSGAKEQILARLAGAALVAKEDAVPVLID-----DALGFT-----DP 831
QY 364 MFAQVTDV-----GMVFV-----PSRAG 382
DB 832 ERLAKMGEVFDITGADGQVIVLTCSPTRYG 861
RESULT 10
GLMS_MYCLE
ID GLMS_MYCLE STANDARD; PRT; 624 AA.
AC P40831;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Glucosamine-fructose-6-phosphate aminotransferase [isomerizing]
DE (EC 2.6.1.16) (Hexosephosphate aminotransferase) (D-fructose-6-
DE phosphate amidotransferase) (G6P) (L-glutamine-D-fructose-6-phosphate
DE amidotransferase) (Glucosamine-6-phosphate synthase).
GN GLMS OR ML0371 OR B229_C3.238.
OS Mycobacterium leprae.
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1769;
RN [1]

RP SEQUENCE FROM N.A.
RA Smith D.R., Robison K.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=TN;
RX MEDLINE=211128732; PubMed=11234002;
RA Cole S.T., Bigmeier K., Parkhill J., James K.D., Thomson N.R.,
RA Wheeler P.R., Honore N., Garnier T., Churcher C., Harris D.,
RA Mungall K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R.M., Devlin K., Duthoy S., Feltham T., Fraser A., Hamlin N.,
RA Holroyd S., Hornsby T., Jagels K., Lacroix C., Maclean J., Moule S.,
RA Murphy L., Oliver K., Quail M.A., Rajandream M.A., Rutherford K.M.,
RA Rutter S., Seeger K., Simon S., Simmonds M., Skelton J., Squares R.,
RA Squares S., Stevens K., Taylor K., Whitehead S., Woodward J.R.,
RA Barrell B.G.;
RT "Massive gene decay in the leprosy bacillus";
RL Nature 409:1007-1011(2001).
CC -I- FUNCTION: Catalyzes the first step in hexosamine metabolism, as a
CC converting fructose-6P into glucosamine-6P using glutamine as a
CC nitrogen source (By similarity).
CC -I- CATALYTIC ACTIVITY: L-glutamine + D-fructose 6-phosphate = L-
CC glutamate + D-glucosamine 6-phosphate.
CC -I- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -I- SIMILARITY: In the C-terminal section; belongs to the SIS family.
CC GFAT subfamily.
CC -I- SIMILARITY: Contains 1 type-2 glutamine amidotransferase domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00020; AAA17307.1; -
DR EMBL; AL583918; CAC29879.1; -
DR PIR; C86955; C86955.
DR PIR; S72993; S72993.
DR HSP; P17169; IMOS.
DR MEROPS; C44.971; -
DR MEROPS; C44. UNW; -
DR Leproma; ML0371; -
DR HAMAP; MF_00164; -; 1.
DR InterPro; IPR000583; GATase_2.
DR InterPro; IPR005855; Glms.
DR InterPro; IPR001347; SIS.
DR Pfam; PF00310; GATase_2; 1.
DR Pfam; PF01380; SIS; 2.
DR TIGR; TIGR01135; glms; 1.
DR PROSITE; PS00443; GATASE_TYPE_II; FALSE NEG.
DR Transferrase; Aminotransferase; Glutamine amidotransferase;
KW Complete proteome.
FT INIT MET 0 0 BY SIMILARITY.
FT ACT_SITE 1 1 GATASE (BY SIMILARITY).
FT ACT_SITE 619 619 ISOMERIZATION FRU-6P (BY SIMILARITY).
FT DOMAIN 1 192 GLUTAMINE AMIDOTRANSFERASE.
FT CONFLICT 554 555 TR -> IS (IN REF. 1).
SQ SEQUENCE 624 AA; 67428 MW; 89D711874D7FD689 CRC64;
Query Match 5.8%; Score 118.5; DB 1; Length 624;
Best Local Similarity 21.4%; Pred. No. 1.8;
Matches 101; Conservative 66; Mismatches 183; Indels 123; Gaps 23;
QY 30 RUTYTPHAARATILAAKAAALSVRDGLGNIIGREGTDPQL-----PATAVGS--- 81
DB 22 RMEYRGYDSSGIALINGSKSNLIVRRRA-----GRLSNLSVLAEMVPAISLAVGVGL 76
QY 82 HFDSVRNGMFGDTAGVVCALAEARVLMENGVNRRHP-----PEFIAIVEEGA- 130
DB 77 HIRWATHGRPTDRNAHPRDGTATKIAVHVINGIIENFPFSLRHEIAGVEFVSDTDEAV 136

QY 280 -VPGEVDFTLRLSPHEESRLVLIDRISVMVGEVASQAGVADVDVEFFNLSP---VQLAP 335
 Db 237 IVCDEHHILAEARSILVPEKQEAQVQKMAAFEAADMGRAEV-EIEVMYGFYQDGD 295
 QY 336 TMDVAVREASALQTHRRDISSGAGHDSMFIAQVTDVGMVFPVSAG--RSHVPEWTFD 393
 Db 296 QVVEITAKAAKIGRPSLEQTSQSGSDANVIA---GHGIPYLVNLAAGVEQIHTKRMPI 352
 QY 394 DDLRKGTEVVLVRMK 408
 Db 353 BELVKTAEVVAIE 367

RESULT 8
 RECN STRCO
 ID RECN STRCO STANDARD; PRT; 572 AA.
 AC Q9S220;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE DNA repair protein recN (Recombination protein N).
 GN RECN OR SC01780 OR SC151.20C.
 OS Streptomyces coelicolor.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Streptomycineae; Streptomycetaceae; Streptomyces.
 OX NCBI_TaxID=1902;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=A3(2) / M145;
 RX MEDLINE=21996410; PubMed=12000953;
 RA Bentley S.D., Chater K.F., Cerdeno-Tarraga A.-M., Challis G.L.,
 RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kiese H.,
 RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,
 RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,
 RA Huang C.-H., Kiese T., Larke L., Murphy L., Oliver K., O'Neill S.,
 RA Rabinowitz E., Rajandream M.A., Rutherford K., Rutter S.,
 RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,
 RA Warren T., Wiatzorrek A., Woodward J., Barrell B.G., Parkhill J.,
 RA Hopwood D.A.;
 RT "Complete genome sequence of the model actinomycete Streptomyces
 RT coelicolor A3(2)";
 RL Nature 417:141-147(2002).
 CC -!- FUNCTION: MAY BE INVOLVED IN RECOMBINATIONAL REPAIR OF DAMAGED
 CC DNA (BY SIMILARITY).
 CC -!- SIMILARITY: Belongs to the recN family.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AL039110; CAB52844.1; -;
 DR PIR; T36883; T36883.
 DR InterPro; IPR003439; ABC transporter.
 DR InterPro; IPR004604; RecN.
 DR TIGRFAME; TIGR00634; recN; 1.
 KW DNA repair; ATP-binding; Complete proteome.
 FT NP_BIND 25 32 ATP (POTENTIAL).
 SQ SEQUENCE 572 AA; 59836 MW; 1D30347D50B8C14 CRC64;

Query Match 6.2%; Score 128; DB 1; Length 572;
 Best Local Similarity 23.7%; Pred. No. 0.39;
 Matches 98; Conservative 58; Mismatches 158; Indels 100; Gaps 21;

QY 11 IKKIRELSRFAEGGVTRLTYPPEAAARETLTAAMKAALSVRDALGNIIGRECT 70
 Db 9 IDDAVVNELS-----PGTAVT---GETGAGTKMTVVT-----SLGLLGR--- 45
 QY 71 DELPAIVAGSHDSVRNGMGFDGTAGVCALEAARVMLENG--YVNRHPFFIAIVEEE 128

Db 46 -ADAAVLRIAGNAVVEGRIAPVGDAAVAVRAEBAGAEELDDGALLISR-----TVSAE 97
 QY 129 GARSSGMLGGRAT-AGLVADRELDLSLVDGVSVRQAATAFGLKPGSLQAARSAAALR 187
 Db 98 GR--SRAHLGGRSVPVGMIAEL-ADLVAVHGQTDQQLKLN---RQQLDRVAGDAV 151
 QY 188 AFIELHIEGQFILBQEQIEIGVVTISVGVRAALRVAVKGRSHAGTTPMHLRQDALV-PAA 246
 Db 152 A-----GPLAKYAEAYRRLRAVVRLEBITTRARERAQEAIDLRLYGLDEIAAVEPRA 203
 QY 247 ---LMVREVNRFVNEIADGTATVGHITVAPGGNQVPGVD-----FTLDLRSP 293
 Db 204 GEDVELAEERLGHAEALASAATVAHAALA--GNPDEPFGVDGATLVAGRAQALDAVRS 261
 QY 294 HEESLRVLIDRI---SVMVGEVASQ-AGVAAADVDEFFNLSPVQLAPTVMVDVAREASALQ 349
 Db 262 HDPALAAEAERIGEVGILLRDRVAGELACYADDLD---ADPLRLA-----AVEERRAALT 312
 QY 350 FTHR-----DISSGAGHDSMFIAQVTDVGMVFPVSAGRSHVPEEWTDDLRKG 399
 Db 313 ALTRYGEDIAAVLSWARQSAARLTTEL-----DGDDEIRIG 347

RESULT 9
 YC78 MYCTU
 ID YC78 MYCTU STANDARD; PRT; 875 AA.
 AC Q11042;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Hypothetical protein RV1278/MT1315/MB1309.
 GN RV1278 OR MT1315 OR MTCV50.04C OR MB1309.
 OS Mycobacterium tuberculosis, and
 OS Mycobacterium bovis.
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773, 1765;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Oeborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.tuberculosis; STRAIN=CDC 1551 / Oshkosh;
 RX MEDLINE=22206494; PubMed=12218036;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouuri H., Gill J., Mikula A.,
 RA Bishai W., Jacobs W.R. Jr., Venter J.C., Fraser C.M.;
 RT "Whole-genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL J. Bacteriol. 184:5479-5490(2002).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC SPECIES=M.bovis; STRAIN=AF2122/97;
 RX MEDLINE=22709107; PubMed=12789872;
 RA Garnier T., Eigmeier K., Camus J.-C., Medina N., Mansoor H.,
 RA Pryor M., Duthoy S., Grondin S., Lacroix C., Monsemp C., Simon S.,
 RA Harris B., Atkin R., Doggett J., Mayes R., Keating L., Wheeler P.R.,
 RA Parkhill J., Barrell B.G., Cole S.T., Gordon S.V., Hewinson R.G.;
 RT "The complete genome sequence of Mycobacterium bovis."

```
DR EMBL; U89279; BAB93857.1; -.
DR PIR; C64783; C64783.
DR EcoGene; EGI3623; allC.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; 1.
KW Hydrolase; Purine metabolism; Complete proteome.
SQ SEQUENCE 411 AA; 45694 MW; C55D1EP854A8F513 CRC64;

Query Match          25.6%; Score 525.5; DB 1; Length 411;
Best Local Similarity 32.7%; Pred. No. 7.7e-27;
Matches 131; Conservative 63; Mismatches 204; Indels 3; Gaps 2;

QY 11 IEKEIRELSRFSAE-GPGVTRLTYTPEHAARETLIAAKMAAALSVRDAGNIIGREG 69
DB 9 IEETLPWSSFGADPAGGMRLLYSPEWLQTOQFKKMAASGLETFDEVGNYLGRNG 68
QY 70 TDELPALAVGSHFDSVRNGMGFDGTAGVVCALAEARVMLENGVYVNRHPPFETIAVEEG 129
DB 69 TEYPPQEVVLGSHIDTVNGGNLDGQFGALAAWLADWLKTOYGCPLRTVEVVMAAEEG 128
QY 130 ARSSGMLGGRATAGLVADELDSLVDGVSVRQAATAFGLXPGELQAAARSADIRAF 189
DB 129 SRFPYFWGSKNIFGLNPDVDVNICDAKNSFVDAMKACGFTLPNAPLTPRO--DIKAF 186
QY 190 IELHIEOGPILEQEQIEIGVTVSIVGVRALRVAVKGRSDHAGTTPMHLRODALVPAALMV 249
DB 187 VELHIEQGCVLENGSQSIGVNAIVGQRRYTVTLNGESNHAGTTPMGYRDRDVTVAFSRIC 246
QY 250 REYNRFVNEIADGTATVATGHLTVAPGGNQVPGVEVFTDLRLSPHEESRLVLIDRISVMV 309
DB 247 HQSVEKAKRGMDPLVLTFGKVEPRNTVNVVPGKTTFTIDCRHTDAAVLRDFTQQLNDM 306
QY 310 GEVASQAGVAADYDEPNLSVQLATPMWDVAVZEAASALQFTHRDSSGAGHDSMFIAQV 369
DB 307 RAICDEMDIGIDILWMDDEEFPVPMKELVATLCELERKLNRYVMHSGAGHDAQIFAPR 366
QY 370 TDVGMVFPVSRAGSRHVPPEWTDFDLRKGTGVVLRVMKAL 410
DB 367 VPTCMFIFPSINGISHNPAERTWITDLAEGVKTLALMLYLQ 407

RESULT 7
YQJE_BACSU
ID YQJE_BACSU STANDARD; PRT; 371 AA.
AC P54542;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Hypothetical protein yqjE.
GN YQJE OR BSU23910.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE=97124195; PubMed=8969508;
RA Mizuno M., Masuda S., Takemaru K.-I., Hosono S., Sato T., Takeuchi M., Kobayashi Y.;
RT "Systematic sequencing of the 283 kb 210 degrees-232 degrees region of the Bacillus subtilis genome containing the skin element and many sporulation genes";
RL Microbiology 142:3103-3111(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S., Borriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A., Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
```

RA Guiseppe G., Guy B.J., Haga K., Haiech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinio S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Meene D., Mellado R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Pardo V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Fumelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone P.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.P., Zumstein E., Yoshikawa H., Danchin A.,
RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
RT *subtilis*,"
RL Nature 390:249-256(1997).
RN [2]
RN FUNCTION.
RC STRAIN=168;
RX MEDLINE=21242727; PubMed=11344136;
RA Schultz A.C., Nygaard P., Saxild H.H.;
RT "Functional analysis of 14 genes that constitute the purine catabolic
RT pathway in *Bacillus subtilis* and evidence for a novel regulon
RT controlled by the PucR transcription activator,"
RL J. Bacteriol. 183:3293-3302(2001).
CC -!- FUNCTION: Involved in the anaerobic utilization of allantoin.
CC Converts allantoin to (S)-ureidoglycolate and ammonia.
CC -!- PATHWAY: Degradation of allantoin (purine catabolism); second
CC step.
CC -!- INDUCTION: Expression is very low in excess nitrogen (glutamate
CC plus ammonia) and is induced during limiting-nitrogen conditions
CC (glutamate). Expression is further induced when allantoin is added
CC during limiting-nitrogen conditions.
CC -!- SIMILARITY: Belongs to peptidase family M40.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; Z99120; CAB15243.1; -.
CC PIR; G70017; G70017.
CC MEROPS; M40.UNW; -.
CC Subtilist; BG13994; pucF.
CC InterPro; IPR002933; Peptidase_M20.
CC Pfam; PF01546; Peptidase_M20; 1.
CC Hydrolase; Purine metabolism; Complete proteome.
CC SEQUENCE 412 AA; 45519 MW; 852BC0952B4B5802 CRC64;

Query Match 29.4%; Score 602.5; DB 1; Length 412;
Best Local Similarity 35.2%; Pred. No. 8.6e-32;
Matches 137; Conservative 70; Mismatches 179; Indels 3; Gaps 2;

QY 22 SAEGPGVRLTTPPEHAAARETLIAAMKAAALSVREDALNIGREGTDPPELPAVGS 81
DB 25 SADG-GVTLTKEMWDLQAVKTEMSSFGLETREDDVGNVFGRLSGTQSDPEIVTGS 83
QY 82 HFDVSVRNGMFGDTAGVCAEAARVYMLENGVYVNRHFFBFAIVEBEGARFSSGMLGGRA 141
DB 84 HDTVINGKYGDAYGVLAAMLALQKLYETGAPKTKTLEAVSLCEEGSRFFPMTYWGSGN 143
QY 142 IAGLVADRLSDLVDDGVSVQQAATAFGKFGELQAARSAADLPAFELHIEQGPIL 201
DB 144 MTGVFSEQAKEPRDESGLQTAMHESFGKGVFQSAVRT--DISAFVELHIEQKTL 201

QY 202 QEQIEIGVVTISGVRLRVAVKGRSDHAGTTPMHLRQDALVPAALMVREYNRFVNEAD 261
DB 202 MSGRDLGIVTSIAGRRVLTLEGECHAGTSMKWRKDPAAASSRIIHELLRSDELDP 261
QY 262 GTVATVGHLTVAPEGNGVQGEVFTDLRSPHEESRLVLDRISSVMVGEVASQGVAAAD 321
DB 262 ELRLTCGKITAEPNVANNVIGRVQVFSIDIRHQHQHVLEPQHQDMVVALINGICLQKGRIV 321
QY 322 VDEFNLSPVQLAPTWDVAEASALQFTHRDLSGGAGHDSMFTAQVTDVGMVVPVPSRA 381
DB 322 IDEYNRIEFPVMDERLKAFAFETALENGFSCEEMVSGAGHDAQMIGRRYPACMLFPVPSRG 381
QY 382 GRSHVPEEWTFDDDLRKGTEVVLVRVMKAL 410
DB 382 GVSHPKEYTSARQLEIGVRALTDLLYLK 410

RESULT 6
ALLC ECOLI
ID ALLC ECOLI STANDARD; PRT; 411 AA.
AC P77425;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Allantoate amidohydrolase (EC 3.5.3.-).
GN ALLC OR GLXB7 OR B0516.
OS Escherichia coli.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Valdes J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12,";
RL Science 277:1453-1474(1997).
RN [2]
RN SEQUENCE FROM N.A.
RA Roberts D., Allen E., Araujo R., Aparicio A., Chung E., Davis K.,
RA Duncan M., Federpspiel N., Hyman R., Kalman S., Komp C., Kurdi O.,
RA Lew H., Lin D., Namath A., Oefner P., Schramm S., Davis R.W.;
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A., AND FUNCTION.
RC STRAIN=K12 / ECL1;
RX MEDLINE=20069628; PubMed=10601204;
RA Cusa E., Obradors N., Baldoma L., Badia J., Aguilar J.;
RT "Genetic analysis of a chromosomal region containing genes required
RT for assimilation of allantoin nitrogen and linked glyoxylate
RT metabolism in *Escherichia coli*,"
RL J. Bacteriol. 181:7479-7484(1999).
CC -!- FUNCTION: Involved in the anaerobic utilization of allantoin.
CC Converts allantoin to (S)-ureidoglycolate and ammonia.
CC -!- PATHWAY: Degradation of allantoin (purine catabolism); second
CC step.
CC -!- INDUCTION: By glyoxylate and allantoin under anaerobic conditions.
CC -!- SIMILARITY: Belongs to peptidase family M40.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AE000157; AAC73618.1; -.
CC EMBL; AB2654; AAB40268.1; -.

```
Query Match          35.1%; Score 719.5; DB 1; Length 409;
Best Local Similarity 38.6%; Pred. No. 2.6e-39;
Matches 155; Conservative 71; Mismatches 165; Indels 11; Gaps 4;

QY 7 QAARIEKIBELRFSAE-GPGVTRLTYPTEHAAARETLIAAKAAALSVDALGNIIIG 65
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 3 QGERLWQRLMELGEVKGQSGGVTRLSFTAERRAKDLVASVMREAGLFVYEDAAGNLIG 62

QY 66 RREGTDPELPAIAGVSHFDSVRNGCMFDGTAGVVCALAAARVMLENGYVNRHFFEFIAIV 125
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 63 RKEGTNPDAFVVLGSLDSVYANGCGDFGLGVLAGVEVVQTMNEHGCVVTHHPIEVVAFT 122

QY 126 EESGARSSGMLGGRATAGIADRELDSDVDEGVSVRQAATAFGLKPGELQAAASAAD 165
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 123 DEEGARFRFGWISRAWAGTLPEALECR-DAEGISLAEMKQAGLDPDRLPOARKPGT 181

QY 186 LRAPIELHIQGPILQEQIEIGVTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPA 245
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 182 VKAAVELHIQGRVLEAGLPVGI VTGIAGLIWVKFTIAGPAEHAGATPMSLRDPMAAA 241

QY 246 ALMV-----RVNRFVNEIADGTATVGHLTVAFGGNGVQVGEVDFTLDLRSPHEESLRVL 301
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 242 AQIIIVIEEARR-----TGTTVTGVLQHVYPGGINVIPERVFEVLDLRLDKAEVRDQV 296

QY 302 IDRISVMVGVSQAQGAADVDEFFNLSPVOLAPTMTVDVAREASALQFTHRIDSSGAGH 361
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 297 WKALAVRAETIAKERNVRLTTERLQEWAPVLCSEVVKQAAERACKQGYGPPFWLPSGAH 356

QY 362 DSMFIAQVTDVGMVFPVPSRAGSHVPEEWTDFDLRLKGTQV 403
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 357 DGVLQAPICPIGMIFVRSQDGVSHSPAENWSTKEDCAVGAEVL 398

RESULT 4
Y588 HAEIN          STANDARD;          PRT;          411 AA.
AC O57051; O05027;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Protein HI0588.
GN HI0588.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;
OC Pasteurellaceae; Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips R., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
CC -!- SIMILARITY: Belongs to peptidase family M40.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
```

```
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by, and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32740; AAC2245.1; -.
DR PIR; D64079; D64079.
DR TIGR; HI0588; -.
DR InterPro; IPR002933; Peptidase_M20.
DR Pfam; PF01546; Peptidase_M20; I.
KW Hydrolase; Complete proteome.
SQ SEQUENCE 411 AA; 45091 MW; CA410B24651525AE CRC64;

Query Match          32.7%; Score 671; DB 1; Length 411;
Best Local Similarity 36.1%; Pred. No. 3.4e-36;
Matches 144; Conservative 80; Mismatches 173; Indels 2; Gaps 2;

QY 10 RIEKEIRELSRFSAE-GPGVTRLTYPTEHAAARETLIAAKAAALSVDALGNIIIGRREG 69
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 7 RVNLEKLAFAISSVPNEULRTLAFTEDDEKAHMIIELCKEYDLSIRRSIGNLFRKAG 66

QY 70 TPELPAIAGVSHFDSVRNGCMFDGTAGVVCALAAARVMLENGYVNRHFFEFIAIVEEEG 129
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 67 KEDFLPAVAGSHDITVNVNAGKDFGPLGSAVAGLEILLQLCEQNIQTRYPLELIIFTEES 126

QY 130 AFSGMLGGRATAGIADRELDSDVDEGVSVRQAATAFGLKPGELQAAASAADLRAF 189
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 127 SRFNFATLGKVMCGIVNQELSSRLDKQKGLSEAMAEVGMFNILVNOAKRDAKEFKCF 186

QY 190 IELHTEQGPILQEQIEIGVTSIVG-VREALRVAVKGRSDHAGTTPMHLRQDALVPAALM 248
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 187 FELHTEQGPRLNEGKTIQVVTGIAAPIRAI-VKIQGADHSGATAMHYRHDLALGSGSEL 245

QY 249 VREVNRFVNEIADGTATVGHLTVAFGGNGVQVGEVDFTLDLRSPHEESLRVLIDRISVM 308
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 246 SLAIERAAIQAGSHVATVGNITAKGVNVMVPGYCELLVDIRGTHVQARDSVFELLOEE 305

QY 309 VGEVASQAQGAADVDEFFNLSPVOLAPTMTVDVAREASALQFTHRIDSSGAGHDSMFIAQ 368
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 306 ISKVSEKRGLLIELQLISKNDPIILLPENVMVQIATAHSLGYSYETIMPSCAGHDAMHMT 365

QY 369 VTDVGMVFPVPSRAGSHVPEEWTDFDLRLKGTQVVLVLM 407
   | : : : | : : : | : : : | : : : | : : : | : : : | : : : | : : : |
Db 366 LCPTGMIFIPSHLGHSHNPLEFTDMKDI EAGIKVLQKVI 404

RESULT 5
ALIC_BACSU
ID ALLC_BACSU          STANDARD;          PRT;          412 AA.
AC O32149;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Allantoate amidohydrolase (EC 3.5.3.-).
GN PUCF OR BSU32530.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunat F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
RA Boriss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruchaci C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emmerson P.T.,
RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
```

Db 121 EVVAFCEEGSRFNDGLFGSRGMVGVKVPEDLQK-VDDNNVTRYBALKTGFGIDPDFTH 179
 Qy 178 AAARSAADLRAPIELHIEOGPILEQIEIGVVTISVGRALRVAVKGRSDHAGTTPMHL 237
 Db 180 QSIRIEGIDKHYFEMHIEOGPILEQIEIGVVTISVGRALRVAVKGRSDHAGTTPMHL 239
 Qy 238 RODALVPAALMVRVNRFP-VNEIADGTATVATGHLTVAPGGGNOVPGVEVDFTDLRSPHBE 296
 Db 240 RKDPLVGAAEVKEVETLGMNDPNATVGTGRIAAFPGGSNIIPESVEFTLDIRDIELE 299
 Qy 297 SLRVLDIRISVMVGEVASOAGVAADVDEFFNLSPLQAPTMVDVAREASALQFTHRDIS 356
 Db 300 RNNKIIKEIKLVNSRTGLEQIEKKNMAAPVVKCSNLSNLSQSKCELEIDAPITV 359
 Qy 357 SGAGHDSFPIAOVTDGMVFPVPSRAGRSHVPEEWTDFDRLRGTEVV 403
 Db 360 SGAGHDMFLABITEIGMVFPVCRNGISHSPKEWAEIDDLTGTKVL 406

RESULT 2

AMB2_BACST
 ID AMB2_BACST STANDARD; PRT; 409 AA.
 AC Q53389;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE N-carbamoyl-L-amino acid hydrolase (EC 3.5.1.87) (L-carbamoylase).
 GN AMAB.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
 ON NCBI_TaxID=1422;
 RX STRAIN=NS1122A;
 RC MEDLINE=94115050; PubMed=7764340;
 RA Mukohara Y., Ishikawa T., Watabe K., Nakamura H.;
 RT "Molecular cloning and sequencing of the gene for a thermostable N-carbamyl-L-amino acid amidohydrolase from Bacillus stearothermophilus strain NS1122A."
 RL Biosci. Biotechnol. Biochem. 57:1935-1937(1993).
 CC -|- CATALYTIC ACTIVITY: N-carbamoyl-L-2-amino acid (a 2-ureido carboxylate) + H(2)O = L-2-amino acid + NH(3) + CO(2).
 CC -|- COFACTOR: Manganese or nickel or cobalt.
 CC -|- SIMILARITY: Belongs to peptidase family M40.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL; S67784; AAC60456.1; -.
 DR PIR; JN0885; JN0885.
 DR MEROPS; M40.UNW; -.
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; 1.
 KW Hydrolase.
 SQ SEQUENCE 409 AA; 44248 MW; 332515C62A1A2F31 CRC64;

Query Match 35.5%; Score 726.5; DB 1; Length 409;
 Best Local Similarity 38.8%; Pred. No. 9.2e-40;
 Matches 156; Conservative 72; Mismatches 163; Indels 11; Gaps 4;

Qy 7 QAARTEKETRELSRFSAE-GPGVTRLTYPTEHAARETLIAAKAAALSVRDALGNITG 65
 Db 3 QGERLWQRLMEVGEVQKPSGSGVTRLSFTAEERRAKDLVASYMREAGLVYEDAAAGNLIG 62
 Qy 66 RREGTDPPELPATVAGSHEDSVNGGMDFGTAGVVCALAEARVYMLGVNRPFFEFIAIV 125
 Db 63 RKEGTNPDPATVTVLVGSHLDSVINGGCFDGLGVLAGVEVVQTNNEHGVVTHHPTEVAPT 122

Qy 126 EEEGARFSSGMLGRATAGLVADRELDSLVDEGVSVRQAATAFGLKPGELQAAARSAD 185
 Db 123 DEEGARFRFGMIGRAMAGTLPPEALECR-DAEGISLAEMAKQAGLDPRLPQAAARKEPT 181
 Qy 186 LRAPIELHIEOGPILEQIEIGVVTISVGRALRVAVKGRSDHAGTTPMHLRQDALVPA 245
 Db 182 VKAYVELHIEOGRVLEETGLPVGIVGTAGLIVKFTIEGKAEHAGATPMSLRRDPMAAA 241
 Qy 246 ALMV-----REYVRFVNEIADGTATVATGHLTVAPGGGNOVPGVEVDFTDLRSPHBEESLRLV 301
 Db 242 AQIIIVIEEARR-----TGTTVTGQHLHYVPGGINVPERVEFVLDRLKAEVRDQV 296
 Qy 302 IDRISVMVGEVASOAGVAADVDEFFNLSPLQAPTMVDVAREASALQFTHRDISSGACH 361
 Db 297 WKATAVRAETIAKERNVRVTTLERLQEMPPVLCSDSEVKRAAEAAACQKLGYPSPWLPSSGAH 356
 Qy 362 DSMFTAOVTDGMVFPVPSRAGRSHVPEEWTDFDRLRGTEVV 403
 Db 357 DSVQLAPICPIGMIFVRSQDGVSHSPAEWSTKEDCAAGAEVL 398

RESULT 3

AMB1_BACST
 ID AMB1_BACST STANDARD; PRT; 409 AA.
 AC P37113; P94345;
 DT 01-OCT-1994 (Rel. 30, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE N-carbamoyl-L-amino acid hydrolase (EC 3.5.1.87) (L-carbamoylase).
 GN AMAB.
 OS Bacillus stearothermophilus.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Geobacillus.
 ON NCBI_TaxID=1422;
 RX STRAIN=NCIB 8224;
 RC MEDLINE=94113715; PubMed=8285691;
 RA Sakanyan V., Desmarz L., Legrain C., Charlier D.R.M., Mett I., Kochikyan A., Savchenko A., Boyen A., Falmagne P., Pifard A., Glansdorff N.;
 RT "Gene cloning, sequence analysis, purification, and characterization of a thermostable aminocyclase from Bacillus stearothermophilus."
 RL Appl. Environ. Microbiol. 59:3878-3888(1993).
 CC -|- CATALYTIC ACTIVITY: N-carbamoyl-L-2-amino acid (a 2-ureido carboxylate) + H(2)O = L-2-amino acid + NH(3) + CO(2).
 CC -|- COFACTOR: Manganese or nickel or cobalt.
 CC -|- SIMILARITY: Belongs to peptidase family M40.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC EMBL; Y08752; CAA69999.1; -.
 DR EMBL; X74289; CAA52341.1; -.
 DR MEROPS; M40.UNW; -.
 DR InterPro; IPR002933; Peptidase_M20.
 DR Pfam; PF01546; Peptidase_M20; 1.
 KW Hydrolase.
 FT CONFLICT 60 62 LIG -> DRE (IN REF. 2).
 SQ SEQUENCE 409 AA; 44166 MW; 9A91F767C9AEFBBB CRC64;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model
Run on: May 3, 2004, 18:14:35 ; Search time 18 Seconds
(without alignments)
1191.828 Million cell updates/sec

Title: US-10-045-063-2
Perfect score: 2049
Sequence: 1 MTLQKAQAARIEKEIRELSR.....FDDLKRGTEWLVKALDR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 141681 seqs, 52070155 residues
Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	727	35.5	414	1 HYUC_PSES	Q01264 pseudomonas
2	726.5	35.5	409	1 AMB2 BACST	Q53389 bacillus st
3	719.5	35.1	409	1 AMB1 BACST	P37113 bacillus st
4	671	32.7	411	1 Y588 HAEIN	Q57051 haemophilus
5	602.5	29.4	412	1 ALLC_BACSU	Q32149 bacillus su
6	525.5	25.6	411	1 ALLC_ECOLI	F77425 escherichia
7	172.5	8.4	371	1 YQJE_BACSU	P54542 bacillus su
8	128	6.2	572	1 RECN_STRCO	Q98220 streptomyce
9	121	5.9	875	1 YC78 MYCTU	Q11042 mycobacteri
10	118.5	5.8	624	1 GLMS MYCLE	P40831 m glucosami
11	117.5	5.7	842	1 VGLH_HSVBC	P27599 bovine herp
12	116	5.7	1085	1 CARE_HSLPU	Q92422 helicobacte
13	114	5.6	627	1 GLMS MYCSM	O68956 m glucosami
14	114	5.6	1102	1 TRAA_RHSIN	P55418 rhizobium s
15	113.5	5.5	3172	1 ER33_SACER	Q03133 saccharopol
16	113	5.5	435	1 NHG3_PSEPU	Q53552 pseudomonas
17	113	5.5	502	1 COBQ_STRCO	Q9rj20 streptomyce
18	111	5.4	415	1 CBPG_PSES6	P06621 pseudomonas
19	110	5.4	513	1 GRDC_EUBAC	P54935 eubacterium
20	110	5.4	1085	1 CARE_HELPY	O25577 helicobacte
21	109.5	5.3	609	1 GLMS VIBPA	Q87sr3 v glucosami
22	108	5.3	521	1 IMH3_CANAL	O00086 candida alb
23	108	5.3	532	1 CH62_VIBPA	Q87j96 vibrio para
24	108	5.3	623	1 GLMS MYCTU	O06253 m glucosami
25	108	5.3	836	1 VG26_BPM2	O64220 mycobacteri
26	107.5	5.2	350	1 HYPE_RHILV	P40599 rhizobium l
27	107.5	5.2	1122	1 RPOB_THECE	P31814 thermococcu
28	107	5.2	432	1 TIG_ECOL6	O8fk47 escherichia
29	106.5	5.2	377	1 DAPF_HAEIN	P44514 haemophilus
30	106.5	5.2	630	1 GIDA_PSEAE	Q9ht09 pseudomonas
31	106	5.2	540	1 CH60_MYCPA	P42384 mycobacteri
32	106	5.2	630	1 GIDA_PSEPU	P25756 pseudomonas
33	106	5.2	631	1 GIDA_PSEPK	Q88rw8 pseudomonas

RESULT 1									
HYUC_PSES									
ID	HYUC_PSES	STANDARD;	PRT;	414 AA.					
AC	Q01264;								
DT	01-OCT-1993 (Rel. 27, Created)								
DT	01-OCT-1993 (Rel. 27, Last sequence update)								
DT	28-FEB-2003 (Rel. 41, Last annotation update)								
DE	Hydantoin utilization protein C (ORF4).								
GN	HYUC								
OS	Pseudomonas sp. (strain NS671).								
OG	Plasmid pHN671.								
OC	Bacteria; Proteobacteria.								
OX	NCBI_TaxID=29441;								
RN	[1]								
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 1-11.								
RX	MEDLINE=92121137; PubMed=1732229;								
RA	Watabe K.;								
RT	"Cloning and sequencing of the genes involved in the conversion of 5-substituted hydantoin to the corresponding L-amino acids from the native plasmid of Pseudomonas sp. strain NS671.";								
RT	J. Bacteriol. 174:962-969(1992).								
RL	-!- FUNCTION: CONVERTS N-CARBAMYL-L-AMINO ACIDS TO L-AMINO ACIDS.								
CC	-!- PATHWAY: CONVERSION OF 5-substituted hydantoin to corresponding L-amino acids.								
CC	-!- SIMILARITY: Belongs to peptidase family M40.								
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).								
CC	EMBL; D10494; BAA01379.1; -.								
DR	EMBL; M72717; AAA25847.1; -.								
DR	MEROPS; M40.UNW. -.								
DR	InterPro; IPR002933; Peptidase_M20.								
DR	Pfam; PF01546; Peptidase_M20; 1.								
KW	Hydrolase; Plasmid; Amino-acid biosynthesis.								
SQ	SEQUENCE 414 AA; 45683 MW; B2DCDE7CEA81977C CRC64;								
Query Match 35.5%; Score 727; DB 1; Length 414;									
Best Local Similarity 37.3%; Pred. No. 8.7e-40;									
Matches 152; Conservative 82; Mismatches 165; Indels 8; Gaps 5;									
QY	1	MTLQKAQAARIEKEIRELSRF-SASGPGVTRITYTPEHAAARETLIAANKAALS	VREDA	59					
Db	4	VTISK---ERLRIHIEQLGEIGKTKDKGVRLSKSDREATLLVSEWNREAGLT	VTVDH	60					
QY	60	LCNIGRREGTPELPATAVGSHPFSGVNGCGFDGTAGVVCALCAARVMLENGYNNRHP	F	119					
Db	61	FNLNIGRKEGETPSPFSGVNGCGFDGTAGVVCALCAARVMLENGYNNRHP	F	120					
QY	120	EFIAIVEEGARFSSGMLGGAIGIAGIADRELDLSLVDGVSVRQA--ATAFGLKPGELQ		177					

O9ew1 streptomyce
O2685 methanobact
O28821 archaeoglob
O9jw10 neisseria m
O05825 drosophila
O8tut7 methanopyru
O9af65 nocardia as
O9a710 caulobacter
P22257 escherichia
Q82e77 streptomyce
O8def3 v glucosami
O9hm66 halobacteri

This Page Blank (uspio)

Job time : 22 secs

```

Db      201 DLENRDVGVTSGYPSGMRVRFSGKTAHTGPTMDLRRNALVAGARWLTAVDGIGWDFP 260
Qy      261 DGT-VATVGHLTVAPEGNGVPGVEVDFTLDRSPHEESLRVLDRISVMVGEVASQAGVA 319
Db      261 VDGKATGARLAAMPNPGILSDTAQAICDVRHPDAATARVMGEMRRRAVFESAARAGCD 320
Qy      320 ADV-DEFFNLSPVQLAPTMVDAREEASALQFTHRDISSGAGHDSMFIAQVTDVGMVFPV 378
Db      321 AVTEDEWFPGDI-FDHAWVEGIRAEAVRMGYDWRDIQSQAGHDAYFLARHCPTAMIFTP 379
Qy      379 SRAGRSHVPEEWTDFDDLKRGTEVYLRVMA 409
Db      380 CKGGITHNEEDCDRDDLAPGLNVLLHAVVA 410

RESULT 15
AE3457
N-carbamoyl-L-amino acid amidohydrolase (EC 3.5.1.1-) [imported] - Brucella melitensis (s
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 15-Feb-2002
C:Accession: AE3457
R;DeiVecchio, V.G.; Kapatal, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
.; Mazur, M.; Goitsman, E.; Selkov, E.; Elzer, P.H.; Hagijs, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A:Title: The genome sequence of the facultative intracellular pathogen Brucella melitens
A:Reference number: AD3252; PMID:11756688
A:Accession: AE3457
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-415 <KUR>
A:Cross-references: GB:AE008917; PIDN:AAL52824.1; PID:g17983664; GSPDB:GNO0190
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMB11643
A:Map position: I
C:Superfamily: N-carbamyl-L-amino acid amidohydrolase
C;Keywords: hydrolase

Query Match      26.2%; Score 536.5; DB 2; Length 415;
Best Local Similarity 32.1%; Pred. No. 1.2e-28;
Matches 132; Conservative 73; Mismatches 177; Indels 29; Gaps 8;

Qy      10 RIEKEIRELSRFSAEFGV---TRLTYTPEHAAARETLIAAKAAALSUREDALGNIIIG 65
Db      13 RLWDSLMEAKI---GGLRGNNRQLTDEDEGERRLFQWCEKAGLSMGVDTMGNNFF 69

Qy      66 RREGTDPELPAIAGVGHFDSVRNGMPDGTAGVVCALAEARVMLENGYVNRHPPEFIAIV 125
Db      70 TRFGEDSDADPVYMGSHLDTPQTGGKFDGVLGGLGVMTLNDNNIRTKRIVVVNWT 129

Qy      126 EEEGARFPSSGMLGGRATAGLVADRELDLSLVDDEGVSVRQAATAFGLKPGELQAAARSAD 185
Db      130 NEEGTRFAPAMLASGVFVGLDQNWAYERTDAAKKTFGELVRIGWK-GDEPVGSRK--- 185

Qy      186 LRAFIEIHTBOGPILBOEQIEIGVTSIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPA 245
Db      186 IHAMFELHTBOGPILBAEHKDIGVTHGQGLWLVLTGTKEAHTGSTPMRMKN-----A 241

Qy      246 ALMVRNVRPNVEIA----DGTVATVGHLTVAPEGNGVPGVEVDFTLDRSPHEESLRVL 301
Db      242 SLGLGKLLQLVNEIAHMAHQPDVAGVGGHIDVSPNSRNVLPQGIVFTVDFRSPNQATLDGM 301

Qy      302 IDRISVMVGEVASQAGVAADVDFNLSVPQLAPTWDVADREASALQFTHRDISSGAGH 361
Db      302 KARFEKAPKIAEELGIGIEIAGHFDPPVTFDTGCVCAIRNAARLGYSHRNIVSGAGH 361

Qy      362 DSMFIAQVTDVGMVFPVPSRAGSH-----VPEEWTFDDLRKCTGVVLRVM 407
Db      362 DACWNRVAPTAMVMCPVDGLSHNEDEDSKEWAS-----AGTDVLLHAV 407
```

Search completed: May 3, 2004, 18:22:30

il, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001
A:Title: Genome sequence of *Yersinia pestis*, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AF0394
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-430 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC92484.1; PID:g15981185; GSPDB:GN00175
C:Genetics:
A:Gene: YPO3249
C:Superfamily: N-carbamyl-L-amino acid amidohydrolase

Query Match 31.4%; Score 642.5; DB 2; Length 430;
Best Local Similarity 36.6%; Pred. No. 9.7e-35;
Matches 145; Conservative 66; Mismatches 172; Indels 13; Gaps 3;

QY 18 LRSFAGGCVTRLTTPPHAAARETLIAAMKAALSUREDALNIGIRREGTDPALPAI 77
DB 24 LAIISSEPLTRVLSPEHLRANRQGVEMQAVGMQVMDTVGNICRYEGRQDPAPAI 83
QY 78 AVGSHFDSVRNGMFDGTAGVCALEAARVMLENGYVNRH-----PREFIAVEEGAR 131
DB 84 LIGSHLDTVRNAGRDGMGVTALEVV-----GYLHRHQRLPVAIEVIGFADDEGTR 137
QY 132 FSSGMLGGRAIAGLVADRELDLSVDEGVSVRQAATAFGLKPGELQAARSADLRAFIE 191
DB 138 FGITLLGSGVTRGWPVEWLNT-TDADGISVAQAVRAGLDPMIDIGQSARAANAFCAYLE 196
QY 192 LHIEQGPILQEQIIEGVVTSIVGVRLAVVAVKGRSDHAGTTPMHLRQDALVPAALMVR 251
DB 197 LHIEQGPILQENAGLGVVTDINGARRLQCFGLAGHAGTTPMGRQDALAGAAEWMCV 256
QY 252 VNRVNEIADGTATVATVGHITVAPGGNOVPGVDEFTDLRSPEESLRVLIDRISVMVGE 311
DB 257 VVALTAQGEHLVATVGTITLCPGAVNVIPGGVRLTDLRGNDRGVNDLLRLAEABA 316
QY 312 VASQAGVAADVDEFFNLSPVQLAPTMVDVRAEASALQFTHRDISSGAGHSMFIAQVTD 371
DB 317 IATRGITPAEAGFYRIKATACDSALQCCISQISQVQRCIALPSGAGHDAIAAEWCWP 376
QY 372 VGMVFPVSRAGSHVPEETDDEDLRKGTGVVLRVM 407
DB 377 VGMVLRCKGGVSHHPDESVTSSDAVAIQAYLEAV 412

RESULT 8
N-carbamoyl-beta-alanine amidohydrolase PA0444 [imported] - *Pseudomonas aeruginosa* (strain ATCC 27852) [GenBank]
C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C83591
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Boman, J.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.N.; Kas, A.; Larbig, K.; Lim, J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.
A:Reference number: AB2950; MUID:20437337; PMID:10984043
A:Accession: C83591
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:Cross-references: GB:AE004481; GB:AE004091; NID:g9946293; PIDN:AAG03833.1; GSPDB:GN00175
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA0444
C:Superfamily: N-carbamyl-L-amino acid amidohydrolase

Query Match 31.2%; Score 639.5; DB 2; Length 427;
Best Local Similarity 38.1%; Pred. No. 1.5e-35;
Matches 156; Conservative 54; Mismatches 190; Indels 9; Gaps 3;

QY 2 TLQQAARIEKEIRLSRFSAGP-GVTRLTYPPEHAARETLIAAMKAALSUREDAL 60

DB 11 TORHIDGQRLWQSLMDLARLGATAGKGGVCRLLSALSDLRQARDLDFQWCEAAGCTVSDRV 70
QY 61 GNIIGRREGTDPALPAVAGSHFDSVRNGMFDGTAGVCALEAARVMLENGYVNRHPE 120
DB 71 GNIIFARRGRNPDLPVVTGSHDITQPTGGKFDGCGVMAGLEVITRLUNDLGVETEAPLE 130
QY 121 FIATVEEGARFSSGMLGGRAIAGLVADRELDLSVDEGVSVRQAATAFGLKPGELQAAA 180
DB 131 VVVWNTNERSRFAFCMGSGVFAKFTLEETLAKRDADGVSGEALDAIG-----YAGA 184
QY 181 RSAA--DURAFIEHIEQGPILQEQIIEGVVTSIVGVRLAVVAVKGRSDHAGTTPMHLR 238
DB 185 RDLCHGPVGAYPEAHIEQGPILQEQIIEGVVTSIVGVRLAVVAVKGRSDHAGTTPMHLR 244
QY 239 QDALVPAALMVRVNRVNFVNEIADGTATVATVGHITVAPGGNOVPGVDEFTDLRSPEESL 298
DB 245 KDALVGAARVAVVNRALAGHOPHACGTGCHAYPGSRNVIPGEGVKTLDLFRHLQPERL 304
QY 299 RVLIDRISVMVGEVASQAGVAADVDEFFNLSPVQLAPTMVDVRAEASALQFTHRDISSG 358
DB 305 DSMIAEVRQVIAATCEKHLQYELVPTADFPPLYPFQDCVGAVEREAAQALGMPQMDIVSG 364
QY 359 AGHDSMFTAQVTDVGMVFPVSRAGSHVPEETDDEDLRKGTGVVLRVM 407
DB 365 AGHDAIFLAELAGPAGMIFVPCENGITSHNEIENASPDLLAAGCAVLLRAM 413

RESULT 9
N-carbamyl-L-amino acid amidohydrolase [imported] - *Caulobacter crescentus* (strain ATCC 31411) [GenBank]
C:Species: *Caulobacter crescentus*
C:Date: 20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 10-May-2001
C:Accession: H87571
R:Nierman, W.C.; Feldblyum, T.V.; Paulsen, I.T.; Nelson, K.E.; Eisen, J.; Heidelberg, J.; B. J.; Laub, M.T.; DeBoy, R.T.; Dodson, R.J.; Durkin, A.S.; Gwinn, M.L.; Haft, D.H.; Kolonitsky, J.; Ermolaeva, M.; White, O.; Salzberg, S.L.; Shapiro, L.; Venter, J.C.; Fraser, C.M.
Proc. Natl. Acad. Sci. U.S.A. 98, 4136-4141, 2001
A:Title: Complete genome sequence of *Caulobacter crescentus*.
A:Reference number: AB7249; MUID:21173698; PMID:11259647
A:Accession: H87571
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-427 <STO>
A:Cross-references: GB:AE005673; NID:g13424176; PIDN:AAK24572.1; GSPDB:GN00148
C:Genetics:
A:Gene: CC2503
C:Superfamily: N-carbamyl-L-amino acid amidohydrolase

Query Match 29.9%; Score 612.5; DB 2; Length 427;
Best Local Similarity 36.1%; Pred. No. 9.9e-34;
Matches 139; Conservative 70; Mismatches 171; Indels 5; Gaps 3;

QY 28 VTRLTYPPEHAARETLIAAMKAALSUREDALNIGIRREGTDPALPAVAGSHFDSVR 87
DB 30 LTRRELTAAHGAALDALAGMAEAGMSARRDAANLIGYEGTHGAKALIGSHIDSVR 89
QY 88 NGGMFDGTAGVCALEAARVMLENGYVNRHPE--EFIAVEEGARFSSGMLGGRAIAGL 145
DB 90 NGRYDGLPGLMGLDGVVLEALHRAQ--RRLPFAIEVAVFGDEEGSRFFASMSCRAIAGT 147
QY 146 VADRELDLSVDEGVSVRQAATAFGLKPGELQAARSADLRAFIEHIEQGPILQEQI 205
DB 148 LDATALE-MKDAEGVSVAELAAFGDPANIASAARRPEEVLAPLEAHIEQGPVLEAEG 206
QY 206 EIGVTSIVGVRLAVVAVKGRSDHAGTTPMHLRQDALVPAALMVRVNRVNEIADGTVA 265
DB 207 ALGVVTAIAAQKRLMVRITGMAGHAGTTFMALRKDPGPAAPAAEAILALERI CRAAGTDLVG 266
QY 266 TVGHLTVAPGGNOVPGVDEFTDLRSPEESLRVLIDRISVMVGEVASQAGVAADVDEP 325
DB 267 TVGRMTALPGAFNVIPGAIEFSDIRTAETSATRDAAVEAITAEIHAIARLDLSATVLM 326

Db 121 EVVAFCEEGRSFNDGLFSGRMVGVKVPEDLQK-VDDNNVTRVEALTKTFGFGIDDPDFT 179
QY 178 AARSAADLRAFLTHIEQGPILBOEQIEIGVTSIVGVRAALRVAVKGRSDHAGTTPMHL 237
Db 180 QSIREIGDIIKHYPFEMHIEQGPYLEKKNYPIGIVSGIAGPSFWKVLVGEAGHAGTVPMSL 239
QY 238 RQDALVPAALMVREVNRF-VNETADGTVATVGHLTVAAPGGNNOVPGVDFDLDLRSRPH 296
Db 240 RKPLVGAAEVKEVETLCHNDPNAPVGTGRIAPFGGSNIIPESVEFLDIRDIELE 299
QY 297 SLRVLIDRISVMVGEVASQAGVAADVDFNFLSPVOLAPTWDVAVREASALQFTHRDIS 356
Db 300 RNNKIIEBKIKLVNTRGLEVQYIEKNMAAPVVKCSENLSLQSKCELEIDAPIIV 359
QY 357 SGAGHDSNFTAQVTDVGMVFPVRAGRSHVPEEWTDFDLDLRKGTGV 403
Db 360 SGAGHDAMFLAEITEIGMVFVRCRNGISHSPKEMAEIDIDILTGKVL 406

RESULT 5
JN0885
N-carbamyl-L-amino acid amidohydrolase (EC 3.5.1.-) [validated] - Bacillus stearothermophilus
C;Species: Bacillus stearothermophilus
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 01-Sep-2000
C;Accession: JN0885; PN0679; I40357
R;Mukohara, Y.; Ishikawa, T.; Watabe, K.; Nakamura, H.
BioSci. Biotechnol. Biochem. 57, 1935-1937, 1993
A;Title: Molecular cloning and sequencing of the gene for a thermostable N-carbamyl-L-amino acid amidohydrolase
A;Reference number: JN0885; MUID:94115050; PMID:7764340
A;Accession: JN0885
A;Molecule type: DNA
A;Residues: 1-409 <MUK>
A;Cross-references: GB:S67784; NID:9460894; PIDN:AAC60456.1; PID:9460895
A;Accession: PN0679
A;Molecule type: protein
A;Residues: 1-13 <M2>
R;Sakanyan, V.; Desmariez, L.; Legrain, C.; Charlier, D.; Mett, T.; Kochikyan, A.; Savchenko, A.; Appl. Environ. Microbiol. 59, 3878-3888, 1993
A;Title: Gene cloning, sequence analysis, purification, and characterization of a thermostable N-carbamyl-L-amino acid amidohydrolase
A;Reference number: I40357; MUID:94113715; PMID:8285691
A;Accession: I40357
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 'DRE', 63-198, 'A', 200-219, 'A', 221, 'P', 223-314, 'L', 316-323, 'A', 325-329, 'EV', 330-331
A;Cross-references: EMBL:X74289; NID:9436795; PIDN:CAA52341.1; PID:9436796
C;Function: catalyzes the hydrolysis of an N-carbamoyl-L-amino acid to ammonia, carbon dioxide and L-amino acid
C;Superfamily: N-carbamyl-L-amino acid amidohydrolase
C;Keywords: hydrolase

Query Match 35.5%; Score 726.5; DB 1; Length 409;
Best Local Similarity 38.8%; Pred. No. 2.1e-41;
Matches 156; Conservative 72; Mismatches 163; Indels 11; Gaps 4;

QY 7 QAARIEKEIRELSRFSNAE-GPGVTRITYTPEHAAARETLIAAKAAALSVDALGNIG 65
Db 3 QGERLWRLMELGEVKQPSGGVTRLSFTAERERRAKDLVASYNREAGLFYEDAGNLIG 62
QY 66 RREGTDPPELPAIAGVSHFSDVRNGMFDGTAGVVCALAEARVNLNGLVYNNRHPFEFIAIV 125
Db 63 RKEGTNPDATVVLGSHLSDVYNGGCFDGLPLVAGVEVVQTNNEHGVTHHPFIEVVAFT 122
QY 126 EEEGARFSSGMLGGRAIAGLVADRELDLSVDEGVSVRQAAATAGFKPGELOAARSAAD 185
Db 123 DEEGARFRFGMIGSRAMAGTLPLPEALECR-DAEGISLAEMKQAGLDPDLPOARPKGT 181
QY 186 LRAFIEHIEQGPILBOEQIEIGVTSIVGVRAALRVAVKGRSDHAGTTPMHLRQDALVPA 245
Db 182 VKAIVSELHIEQGPVLETSGLFVGIIVGIAGLIWKFTIEGKAHAGATPMSLRDRDMAA 241
QY 246 ALMV-----RVNRFVNEIADGTVATVGHLTVAAPGGNNOVPGVDFDLDLRSRPH 301

Db 242 AQIIIVIEBEARR-----TGTTVGTVQLHVYPGGINVIPERVEFLDLRLDKAEVRDQV 296
QY 302 IDRISVMVGEVASQAGVAADVDFNFLSPVOLAPTWDVAVREASALQFTHRISSAGH 361
Db 297 WKAIARAETIAKERNVRVTTERLQEMPVPLCSDEVKRAABAAACQKLGYPFSLPSSAAH 356
QY 362 DSMFIAQVTDVGMVFPVRAGRSHVPEEWTDFDLDLRKGTGV 403
Db 357 DSVQLAPICPIGMIFVRSQGVSHSPAENSTKEDCAAGAEVL 398

RESULT 6
D64079
probable N-carbamyl-L-amino acid amidohydrolase (EC 3.5.1.-) HI0588 [similarity] - Haemophilus influenzae
C;Species: Haemophilus influenzae
C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 08-Sep-2000
C;Accession: D64079
R;Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.R.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M. Science 269, 496-512, 1995
A;Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A;Reference number: A64000; MUID:95350630; PMID:7542800
A;Accession: D64079
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-411 <YIGR>
A;Cross-references: GB:U32740; GB:L42023; NID:g1573572; PIDN:AAC22245.1; PID:g1573578;
C;Superfamily: N-carbamyl-L-amino acid amidohydrolase
C;Keywords: hydrolase

Query Match 32.7%; Score 671; DB 1; Length 411;
Best Local Similarity 36.1%; Pred. No. 1.1e-37;
Matches 144; Conservative 80; Mismatches 173; Indels 2; Gaps 2;

QY 10 RIEKEIRELSRFSNAE-GPGVTRITYTPEHAAARETLIAAKAAALSVDALGNIGRREG 69
Db 7 RVONLIEKLAFTISSVPNEUTRLAFTDEDEKAHMIIELCKEYDLSIRRSIGMLFIRKAG 66
QY 70 TDPPELPAIAGVSHFSDVRNGMFDGTAGVVCALAEARVNLNGLVYNNRHPFEFIAIV 129
Db 67 KEDFLPAVAFGSHDITVNVNAGKFDGPLGSLVAGLEILLQCEQNIQTRYPLELIIFTCEES 126
QY 130 APFSSGMLGGRAIAGLVADRELDLSVDEGVSVRQAAATAGFKPGELOAARSAADLRAF 189
Db 127 SRFNFTATLGSVKVCGIVNQEKLSSLRDKQKGLSEAMVGMFNILVNOAKRDAKEFKCF 186
QY 190 IELHIEQGPILBOEQIEIGVTSIVG-VRALRVAVKGRSDHAGTTPMHLRQDALVPAALM 248
Db 187 FELHIEQGPRLNEGKTIGVVTGIAAPIRAI-VKIQADHSGATAMHYRHDALLGGSSEL 245
QY 249 REVNRFVNEIADGTVATVGHLTVAAPGGNNOVPGVDFDLDLRSRPHESLRVLIDRISVM 308
Db 246 SLAIERAAIQAGHSTVATVGNITAKPGVMNVVPGYCELLVDIRGTHVQVARDVSFELLQEE 305
QY 309 VGEVASQAGVAADVDFNFLSPVOLAPTWDVAVREASALQFTHRISSAGHDSNFIAQ 368
Db 306 ISKVSEKRGLLLEQLQISKDNPIILPNMVMNIAETAHSLGYSYEIMPSSGAGHDAMHMT 365
QY 369 VTDVGMVFPVRAGRSHVPEEWTDFDLDLRKGTGVLRVM 407
Db 366 LCPTGMIFIPSHLGISHNPLEFTDWKDEAGIKVLQKVI 404

RESULT 7
AH0394
probable amino acid hydrolase YPO3249 [imported] - Yersinia pestis (strain CO92)
C;Species: Yersinia pestis
C;Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 09-Nov-2001
C;Accession: AH0394
R;Parkhill, J.; Wren, B.W.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Tarraga, A.M.; Chillingworth, T.; Cronin, R.M.; Davies, P.; Dougan, G.;

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 18:18:20 ; Search time 20 Seconds
(without alignments)
1981.545 Million cell updates/sec

Title: US-10-045-063-2
Perfect score: 2049
Sequence: 1 MTLQKAQARIKEIRELSR.....FDDLKRGTEVLRVMAKLDLR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 78:**

1: Piri:**

2: Piri2:**

3: Piri3:**

4: Piri4:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	858	41.9	423	2 AB1142	N-carbamyl-L-amino
2	853	41.6	414	2 AE1500	N-carbamyl-L-amino
3	762.5	37.2	414	2 AE3745	N-carbamyl-L-amino
4	728	35.5	414	1 D42594	N-carbamyl-L-amino
5	726.5	35.5	409	1 JN0885	N-carbamyl-L-amino
6	671	32.7	411	1 D64079	probable N-carbamyl
7	642.5	31.4	430	2 AH0394	probable amino aci
8	639.5	31.2	427	2 C83591	N-carbamoyl-beta-a
9	612.5	29.9	427	2 H87571	N-carbamyl-L-amino
10	602.5	29.4	412	1 G70017	probable N-carbamyl
11	562	27.4	409	2 AG2869	N-carbamoyl-beta-a
12	562	27.4	415	2 A97846	N-carbamoyl-beta-a
13	562	27.4	416	2 F75429	N-carbamyl-L-amino
14	554.5	27.1	414	2 H95999	probable N-carbamyl
15	536.5	26.2	415	2 AE3457	N-carbamoyl-L-amino
16	534.5	26.1	411	2 B90701	allantoate amidohy
17	534.5	26.1	411	2 E85551	probable hydantoin
18	528.5	25.8	411	2 A10567	allantoate amidohy
19	528.5	25.8	425	2 F85227	hyuC-like protein
20	525.5	25.6	411	1 C64783	probable N-carbamyl
21	525	25.6	401	2 F90446	N-carbamoyl-L-amino
22	501	24.5	413	2 AH2933	N-carbamoyl-beta-a
23	501	24.5	441	2 G98348	N-carbamoyl-beta-a
24	466.5	22.8	338	2 T04891	hyuC protein homol
25	172.5	8.4	371	2 E69963	tripeptidase homol
26	171	8.3	400	2 B72650	hypothetical prote
27	168	8.2	372	2 E83833	hypothetical prote
28	148	7.2	412	2 F87624	peptidase, M20/M25
29	146.5	7.1	381	2 E75298	probable acetylor

30	144	7.0	354	2 E71193	probable operon pr
31	139	6.8	346	2 D75152	aminopeptidase m42
32	139	6.8	805	2 G87662	peptidase, M20/M25
33	135.5	6.6	377	2 H89930	hypothetical prote
34	133.5	6.5	108	2 T05921	N-carbamyl-L-amino
35	133.5	6.5	412	2 C83297	carboxypeptidase G
36	131.5	6.4	368	2 B82211	peptidase, M20A fa
37	130.5	6.4	377	2 C82113	succinyl-diaminopi
38	128	6.2	383	2 G83500	succinyl-diaminopi
39	128	6.2	572	2 T36883	probable DNA repai
40	127	6.2	443	2 T35974	probable aminoacyl
41	127	6.2	1214	2 AG2897	conserved hypothet
42	127	6.2	1387	2 A97673	probable periplasm
43	126	6.1	353	2 E71029	probable Frv opero
44	126	6.1	353	2 H84303	endoglucanase [imp
45	126	6.1	643	2 H84305	Htr8 transducer [1

ALIGNMENTS

RESULT 1

AB1142

N-carbamyl-L-amino acid amidohydrolase homolog lmo0537 [imported] - Listeria monocytoge

C;Species: Listeria monocytogenes

C;Date: 27-Nov-2001 #sequence_revision 27-Nov-2001 #text_change 14-Dec-2001

C;Accession: AB1142

R;Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloecke

.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Duesurget, O.; Entian, K.D.; Fsihi, H.

D.; Jones, L.M.; Karst, U.

Science 294, 849-852, 2001

A;Authors: Kreft, J.; Kuhn, M.; Kunst, F.; Kurapkat, G.; Madueno, E.; Maitournam, A.; M.

ok, C.; Schlueter, T.; Simoes, N.; Tierrez, A.; Vazquez-Boland, J.A.; Voss, H.; Wehland

A;Title: Comparative Genomics of Listeria species.

A;Reference number: AB1077; MUID:21537279; PMID:11679669

A;Accession: AB1142

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-423 <GLA>

A;Cross-references: GB:NC_003210; PIDN:CAC98616.1; PID:gl6409913; GSPDB:GN00177

A;Experimental source: strain EGD-e

C;Genetics:

A;Gene: lmo0537

C;Superfamily: N-carbamyl-L-amino acid amidohydrolase

Query Match 41.9%; Score 858; DB 2; Length 423;

Best Local Similarity 44.2%; Pred. No. 3.3e-50;

Matches 177; Conservative 68; Mismatches 153; Indels 2; Gaps 2;

QY	10	RIEKEIRELSRSA-EGPGVTRLTTPPEHAAARETLIAAMKAAALSVREDALNIIGRRE	68
DB	7	RKKHLEKUDTFTATPGQGTTRLTYSKEDLDARNYLKQEMAKVGLTVSEDAIGNIYGRLE	66
QY	69	GTDPELPATAVGSHFDSVRNGMFDGTAGVVCALAAARVMLENGYVNRPFETIAVEEE	128
DB	67	GSPDLPVAVGSHFDSVRNGMFDGTAGVVCALAAARVMLENGYVNRPFETIAVEEE	126
QY	129	GARFSSGMIGGRAIAGLVADRELDLSLVBEGVSVROAATAFGLKPGELQAARSADLRA	188
DB	127	GSRFGAGLLASRAITGKVTTEMLHEMKDIDGITATDAMAKLGFNANVRHRAITKESVKA	186
QY	189	FIELHIEQGPILLEQEQIEIGVVTISIVGVPRALVAVKGRSDHAGTTPMHLRODALVPAALM	248
DB	187	FIELHIEQGPILLEQEQIEIGVVTISIVGVPRALVAVKGRSDHAGTTPMHLRODALVPAALM	246
QY	249	VREVNRFVNEIADGTATVATVGHILTAPCGGNQVPGVDFTLDRSPHEESLRVLIDRISVM	308
DB	247	LQGLPELAIQEGGGTTLTGKLVNYPNGANVPDKVIFTVDFRKAKEIHVQNTLAKTKEI	306
QY	309	VEVEASQAGVADVDEFFNLSVPQLAPTWTVDVAREAASALQFTHRDISGAGHDSMFIAQ	368
DB	307	I-QATEKNGITCEIKMDIYQQPETHLSKEIHAQALTESADQLGFKYRTWVSAGHDAMIFAS	365

Db 9 ELLPTGRSSASG-GYRRFAWTCGADACDCAWFKQAESRGLTYELDRNGNQWAWLG----- 62
QY 71 DPGL-PAIATVAGSHFDSVRNGMGFDGTAGVVCALAEARVNLGVNRRHPFFEFIAIVEBEG 129
Db 63 DPTAGDVVTVGSHLDSVPDGGAFDGLGVVSSFAALDELARQVTLAKPLALVNFGEDEG 122
QY 130 ARFSSGMLGGRIAGLIVADRELDLSVDEGVSVROAATAFGLKPGELQAAARSADLRAF 189
Db 123 ARFLGACVGSRLAAGOLTVQAHRLTDADGVTLFQAMERAGYDAGTIGDPDPERLARIGAF 182
QY 190 IELHTEQGPILBOEQIEIGVVTISIVCVRALRVAVKGRSDHAGCTPMHLRQDALVPAALMV 249
Db 183 VELHVEQGRALDSDGRIGIASAIWPHGWRDFRGEANHAGTTLAORRDPMLSVETV 242
QY 250 REVNRFVNEIADGTVATVAGGNGQVCEVDFTLDRSPHEESLRVLIDRISVMV 309
Db 243 LAARREAR--LAGAVATFKIAVEPNGVNAISLVGRWLDSRAADQRTLDTVTVGKAA 300
QY 310 GEVASOAGVAADVDFEFLSPVLAPTWDVDAVREA-ASALQTH-----RD--- 354
Db 301 REYADAHGTELDV-----VRESFTPVVEFDHALRDELARILGRDGTG 342
QY 355 -----ISSGAGHDSMFIAQVTDVGMVFPVPSRAGRSHVPEEWTDFDRLKRGTEV 402
Db 343 GAGTGADLKVPIVGTGAGHAGDILSGTPTAMLFVRNPTGVSHSPAEEAEDDCVAGVTA 402
QY 403 VLVRVMKALDR 412
Db 403 LADVLEGLAR 412

RESULT 12
US-10-282-122A-47606
; Sequence 47606, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47606
; LENGTH: 319
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47606
Query Match 7.0%; Score 143.5; DB 12; Length 319;
Best Local Similarity 27.5%; Pred. No. 5.8e-05;
Matches 44; Conservative 19; Mismatches 66; Indels 31; Gaps 4;
QY 3 LKQQAARIEKIEBELSRFSAGCGVCTRLTTPPEHAARETLIAMKAAALSVRDAL-- 60
Db 39 VQLOQASIVGTITSL-----GFTVRYTTTSHGVAASDWLALQWKLAGARADITVE 91
QY 61 -----GNIGRREGTDPALPAIAGVSHFDS-----VRNGMGFDGTAGVVCAL 103
Db 92 QVHTGFPQKSVILTRGSDPAAGTVVLGGHLDSTVGRTTENTRSFGADDDASGLASLT 151
QY 104 AARVNLGVNRRHPFFEFIAIVEBEGARFSSGMLGGRAIA 143
Db 152 ALRVLLANNYRPKRTIKFVGAAEE-----AGLLGSKAIA 186
RESULT 13
US-10-282-122A-50358
; Sequence 50358, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 50358
; LENGTH: 418
; TYPE: PRT
; ORGANISM: Burkholderia mallei

```

US-10-369-493-10895
; Sequence 10895, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 10895
; LENGTH: 380
; TYPE: PRT
; ORGANISM: Ferroplasma acidarmanus
US-10-369-493-10895

Query Match      23.3%; Score 476.5; DB 15; Length 380;
Best Local Similarity 31.9%; Pred. No. 1.8e-36;
Matches 123; Conservative 57; Mismatches 130; Indels 15; Gaps 6;

Qy 23 AEGPGVTRLTYTPEHAAARETLIAAKAAALSVDALGNIGR-REGTDPPELPAIAVGS 81
Db 7 AEENGVRNALNEVDIQARKNLEGEEMRAVKADIKHDDAGLIFGTLGSGKD----NTAIGS 62

Qy 82 HFDVSRNGMGFDGTAGVVCALAEARVM---LENGYVNRHPPEFTIAVEEERGAFSSGMLG 138
Db 63 HMDVSPNGRFDGFGYGVMSGMLLKDLGSTLKNRKITA-----IDFTNSEGARFQPSLLG 117

Qy 139 GRAIAGLVADRELDLSVDEGVSVRQAATAFGLKPGELQAAARSADLRAFIELHIEQGP 198
Db 118 SGMSTGVFTKEFAYSRKDSGITFEALEKSGYM-GEEQNRVKN-QDITRYLELHIEQGP 175

Qy 199 ILEOQIEIGVVTISVGVRLRVAVKGRSDHAGTTPMHLRQDALVPAALMVREVRVNE 258
Db 176 VLEENYQIGPKGIVTMVDNISFTGESNQAGPTPMKVRKDALVAASRFTVAVRDMAKS 235

Qy 259 IADGTAVTGHILTVAAPGCGNQVPEVFTLDLRSPHEESLRVLIDRISVMVGEVASOAGV 318
Db 236 SGKELTWTGKINNFPNAYNVPKVTMNLDRSPDKKTAEEYSKMAKISEKIASSEGV 295

Qy 319 AADVDFEFLNLPVLAPTMDVADVAEASALQFTHRDISSGAGHDSMFIAQVTDVGMVFVP 378
Db 296 PVDFKQWTTETTFDEMRKEIKKSCSLNLYKELYSWPGHDAQYMNVRVPTAMIFIP 355

Qy 379 SRAGSRHVPPEWTFDDLRKGTGV 403
Db 356 SHNGRSHTKEEYSSDRDLINGSVL 380

RESULT 10
US-10-369-493-11009
; Sequence 11009, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; PRIOR FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 11009
; LENGTH: 413
; TYPE: PRT
; ORGANISM: Streptomyces avermitilis
US-10-156-761-11032

Query Match      20.0%; Score 409.5; DB 14; Length 413;
Best Local Similarity 29.1%; Pred. No. 4.7e-30;
Matches 125; Conservative 56; Mismatches 192; Indels 57; Gaps 9;

Qy 14 EIRELSRFAEGPGVTRLTYTPEHAAARETLIAAKAAALSVDALGN---IIGRREGT 70

```

Db 60 ELGKVSASGYLERTFLSPAFMXAIDLIRKWMEDAGLRTQEXMGVHGLVDGANGNAEA 119
QY 77 IAVGSHFDSVRNGMGFDGTAGVVCALAAARVMLENGYVN--RHPFEFFIAVEEGARFSS 134
Db 120 VXLGSHMDTVGNAGMFDGSLGIVSAISAKAMHVNGKQKLRPVEVIAFSDDEEGRFQT 179
QY 135 GMLGGRATAGLVADREDSLVDEGVSVRQAATAFGLKPGELQAARSADLR----- 187
Db 180 TFLGSAIAGILPGTTLE-ISKREVMIKDF-----LKENSMDITEESLLKLYDPKSIW 233
QY 188 AFIELHIEQGPILQEQIBIGVVTISVGRALRVAVKGRSDHAGTTPMHLRQDALVPA-- 246
Db 234 GYVEVHIEQGPVLEQVGFPLGVKGIAGQTRLKVTVRGQGHAGTVPMSMRQDPMAAAE 293
QY 247 -LMVRE-----VNRFNIEADGTATVATGHLTVAPGGGNQVPEVDFTLD 289
Db 294 QIVVLESCKHPEEYLSYDCHCSDSTVKSLSLVCVTGVEISTWPSASNVIPGQVTVVD 353
QY 290 LRSPEESLRVLIDRISVMVGEVASQAGVAADVDFNLSVQLAPTVMVDVREAA--SAL 348
Db 354 IRAIDDLGREAVIYDLISKIYQICDKRSVSCIIEHKHDAGAVICDLSLSQLKSAAYSAL 413
QY 349 QFTHRDIS-----SGAGHDSMFIAQVTDVGMVFPVPSRAGRSHVPEE 389
Db 414 KXMEGDIQDEVPTLMSGAGHDAMAISHLTKEGMLFVRCRGGISHSPQE 461

RESULT 7

US-10-425-114-43946
; Sequence 43946, Application US/10425114
; Publication No. US2004003488A1

GENERAL INFORMATION:

; APPLICANT: Liu, Jingdong

; APPLICANT: Zhou, Yihua

; APPLICANT: Kovalic, David K.

; APPLICANT: Screen, Steven E

; APPLICANT: Tabaska, Jack E

; APPLICANT: Cao, Yongwei

; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53313)B

; CURRENT APPLICATION NUMBER: US/10/425,114

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 73128

; SEQ ID NO 43946

; LENGTH: 511

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; OTHER INFORMATION: Clone ID: 700682144_FLI.ppe

US-10-425-114-43946

Query Match 25.5%; Score 522.5; DB 12; Length 511;
Best Local Similarity 32.1%; Pred. No. 1.2e-40;
Matches 131; Conservative 73; Mismatches 163; Indels 41; Gaps 8;

QY 17 ELSRFAEGPGVTRLTYPPEHAAARETLIAAMKAALSVRDALGNIIGRGTGDPALPA 76
Db 88 ELGKVSASGYLERTFLSPAFMXAIDLIRKWMEDAGLRTQEXMGVHGLVDGANGNAEA 147
QY 77 IAVGSHFDSVRNGMGFDGTAGVVCALAAARVMLENGYVN--RHPFEFFIAVEEGARFSS 134
Db 148 LLIGSHMDTVGNAGMFDGSLGIVSAISAKAMHVNGKQKLRPVEVIAFSDDEEGRFQT 207
QY 135 GMLGGRATAGLVADREDSLVDEGVSVRQAATAFGLKPGELQAARSADLR----- 187
Db 208 TFLGSAIAGILPGTTLE-ISKREVMIKDF-----LKENSMDITEESLLKLYDPKSIW 261
QY 188 AFIELHIEQGPILQEQIBIGVVTISVGRALRVAVKGRSDHAGTTPMHLRQDALVPA-- 246
Db 262 GYVEVHIEQGPVLEQVGFPLGVKGIAGQTRLKVTVRGQGHAGTVPMSMRQDPMAAAE 321
QY 247 -LMVRE-----VNRFNIEADGTATVATGHLTVAPGGGNQVPEVDFTLD 289

Db 322 QIVVLESCKHPEEYLSYDCHCSDSTVKSLSLVCVTGVEISTWPSASNVIPGQVTVVD 381
QY 290 LRSPEESLRVLIDRISVMVGEVASQAGVAADVDFNLSVQLAPTVMVDVREAA--SAL 348
Db 382 IRAIDDLGREAVIYDLISKIYQICDKRSVSCIIEHKHDAGAVICDLSLSQLKSAAYSAL 441
QY 349 QFTHRDIS-----SGAGHDSMFIAQVTDVGMVFPVPSRAGRSHVPEE 389
Db 442 KXMEGDIQDEVPTLMSGAGHDAMAISHLTKEGMLFVRCRGGISHSPQE 489

RESULT 8

US-10-424-599-229957
; Sequence 229957, Application US/10424599
; Publication No. US20040031072A1

GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 229957

; LENGTH: 448

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: unsure

; LOCATION: (1)-(448)

; OTHER INFORMATION: unsure at all Xaa locations

; FEATURE:

; OTHER INFORMATION: Clone ID: PAT_MRT3847_49676C.1.ppe

US-10-424-599-229957

Query Match 24.6%; Score 504; DB 12; Length 448;
Best Local Similarity 34.1%; Pred. No. 5.7e-39;
Matches 135; Conservative 57; Mismatches 178; Indels 26; Gaps 8;

QY 8 AARIEKEIRELSRFS-AEGPGVTRLTYPPEHAAARETLIAAMKAALSVRDALGNIIGR 66
Db 53 AALQNLQIDELAAFSDSPAPSVTVLYTQDVLGRRYVKTQWELAGLSVREDAGVGNIFGR 112
QY 67 REGTDPPELPAIYAVGSHFDSVRNGMGFDGTAGVVCALAAARVMLENGYVNRHPPFEFIAIVE 126
Db 113 WDGLEPELAAVATGSHIDAIPYSGKYDGVGVGLGAIEAIRVLKRSFGKPRRPLEVSETS 172
QY 127 EEGARFSSGMLGGRAIAGLV-ADRELDLSLVDEGVSVRQAATAFGLKPGKE--LQAARSA 183
Db 173 EEPTRFGICGLGRLLAGSEDLANSKLTSTDIQNIISFLDAAGSAGYSKNEDDLSSVFLKK 232
QY 184 ADLAFIELHIEQGPILQEQIBIGVVTISVGRALRVAVKGRSDHAGTTPMHLRQDALV 243
Db 233 GTYSFAVELHIEQGPILQEQIBIGVVTISVGRALRVAVKGRSDHAGTTPMHLRQDALV 292
QY 244 PAALMVREVRNVF-NEIADG---TVATVGHULTVAPGGGNQVPEVDFTLDLRSPEESLR 299
Db 293 AASELALAVRHLVDSVPISGTDVFTVDCLEHLSGAINSIPSKYHLEIDTRDIDEERN 352
QY 300 VLIDRISVMVGEVASQAGVAADVDFNLSVQLAPTVMVDVREAAASALQFTHRDISSA 359
Db 353 KVEKIHQSAIKITKTRGVKLSDFRVINQDP-----PAIYD---BEVFKLAET----- 397
QY 360 GHDSMFIAQVTDVGMVFPVPSRAGRSHVPEEWTDFDD 395
Db 398 ---TXYMARTSPMSMFFIPCFKRYSLKPECFATIED 430

RESULT 9

```
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 49502
; LENGTH: 472
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: 701134260_FLI.pep
US-10-425-114-49502
```

```
Query Match 30.6%; Score 628; DB 12; Length 472;
Best Local Similarity 36.4%; Pred. No. 1e-50;
Matches 149; Conservative 66; Mismatches 188; Indels 6; Gaps 5;
```

```
QY 8 AARIEKEIRELSRFS-AEGPGVTRLTYTPEHAAARETLIAAKKAALSVRDAGLNIIGR 66
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 AQAALQNQIDELAAAFSDSPAPSVTRVLYTDKVLGRRYVKTOMELAGLSVREDVAGNIFGR 120
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 67 REGTDPALPAIVGSHFSDSVNRGGMFDGTAGVVCALAEAAARVMLENGYVNRHPPEFFIAIVE 126
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 121 WDLGLEPAAVATGSHIDAIPYSGKYDGVGVGLGATEAIRVLKRSFGKPRRPLEVISFTS 180
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 127 EGARFSSGMLGGRATAGLV-ADRELDLSLDEGVSVRQAAATAFGLKPGE--LQAAARSA 183
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 181 EEPTRFGICGLSRLLAGSDDLANSKLTSTDIQNISFLDAAGSAGYSKNEDDLSSVFLKK 240
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 184 ADLRAFIEHIEOGPILEQIEQIEIGVVTISIVGRALRVAVKGRSDHAGTTPMHLRQDALV 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 241 GTYSAPVELHIEOGPILEDEGISIGIVTAIAAPASLTVEFEGNGHAGAALMPNRNDAGL 300
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 PAALMVRENVRFNEIAD-GTVATVGHLTVPAGGGNQVPGVEVDTTDLRSPHEESLRVLI 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 301 AASELALAVRHLVDSGSDITVGTGILELHPGAINSIKSHLEIDTRDIDEERNKVV 360
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 303 DRISVMVGEVASOAGVAADVDEFNLSPVQLA-PTWMDAVREAASALQFTHRDISSGAGH 361
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 361 EKIHQSAIKITKTRGVKLSDFRVINQDPPAISDEAVIKAVETATKELNLTSLKLMISRAVH 420
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 362 DSMFIAQITVDGVMFVPSRAGRSHVPEEWTDFDRLKRGTEVVLVRVMKAL 410
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 421 DSLFMARLSPMGIMFIPCYKGYSHKPEEFATIEDMSNGVKVLAALTAKL 469
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
RESULT 5
US-10-424-599-229959
; Sequence 229959, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 229959
; LENGTH: 460
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
```

```
; NAME/KEY: unsure
; LOCATION: (1)..(460)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_49678C.1.pep
US-10-424-599-229959
```

```
Query Match 30.5%; Score 625; DB 12; Length 460;
Best Local Similarity 36.4%; Pred. No. 1.9e-50;
Matches 149; Conservative 65; Mismatches 189; Indels 6; Gaps 5;
```

```
QY 8 AARIEKEIRELSRFS-AEGPGVTRLTYTPEHAAARETLIAAKKAALSVRDAGLNIIGR 66
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 49 AQAALQNQIDELAAAFSDSPAPSVTRVLYTDKVLGRRYVKTOMELAGLSVREDVAGNIFGR 108
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 67 REGTDPALPAIVGSHFSDSVNRGGMFDGTAGVVCALAEAAARVMLENGYVNRHPPEFFIAIVE 126
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 109 WDLGLEPAAVATGSHIDAIPYSGKYDGVGVGLGATEAIRVLKRSFGKPRRPLEVISFTS 168
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 127 EGARFSSGMLGGRATAGLV-ADRELDLSLDEGVSVRQAAATAFGLKPGE--LQAAARSA 183
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 169 EEPTRFGICGLSRLLAGSDDLANSKLTSTDIQNISFLDAAGSAGYSKNEDDLSSVFLKK 228
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 184 ADLRAFIEHIEOGPILEQIEQIEIGVVTISIVGRALRVAVKGRSDHAGTTPMHLRQDALV 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 229 GTYSAPVELHIEOGPILEDEGISIGIVTAIAAPASLTVEFEGNGHAGAALMPNRNDAGL 288
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 PAALMVRENVRFNEIAD-GTVATVGHLTVPAGGGNQVPGVEVDTTDLRSPHEESLRVLI 302
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 289 AASELALAVRHLVDSGSDITVGTGILELHPGAINSIKSHLEIDTRDIDEERNKVV 348
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 303 DRISVMVGEVASOAGVAADVDEFNLSPVQLA-PTWMDAVREAASALQFTHRDISSGAGH 361
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 349 EKIHQSAIKITKTRGVKLSDFRVINQDPPAISDEAVIKAVETATKELNLTSLKLMISRAVH 408
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 362 DSMFIAQITVDGVMFVPSRAGRSHVPEEWTDFDRLKRGTEVVLVRVMKAL 410
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 409 DSLFMARLSPMGIMFIPCYKGYSHKPEEFATIEDMSNGVKVLAALTAKL 457
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

```
RESULT 6
US-10-424-599-158434
; Sequence 158434, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 158434
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(483)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_114084C.1.pep
US-10-424-599-158434
```

```
Query Match 25.7%; Score 526.5; DB 12; Length 483;
Best Local Similarity 32.1%; Pred. No. 4.7e-41;
Matches 131; Conservative 74; Mismatches 162; Indels 41; Gaps 8;
```

```
QY 17 ELSRFSAREPGVTRLTYTPEHAAARETLIAAKKAALSVRDAGLNIIGRREGTDPALPA 76
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```

Qy	121	FTATVEE	GARSSG	MLG	RAIAGL	VADRELD	LSLVDE	DGVS	VRQAAT	AFGLK	GPGLQAAA	180
Db	121	FTATVEE	GARSSG	MLG	RAIAGL	VADRELD	LSLVDE	DGVS	VRQAAT	AFGLK	GPGLQAAA	180
Qy	181	RSAADL	RAPIEL	HI	EQGP	ILEEQ	PIEQE	QIEI	GVVTSI	VGVRAL	RVAVKGRSDHAGTT	PMHLROD 240
Db	181	RSAADL	RAPIEL	HI	EQGP	ILEEQ	PIEQE	QIEI	GVVTSI	VGVRAL	RVAVKGRSDHAGTT	PMHLROD 240
Qy	241	ALVPAAL	MYREVN	RFNEI	ADGT	VATV	GHLTV	APGCGG	NOVP	GEVD	FTLDL	RSRPHESLSRV 300
Db	241	ALVPAAL	MYREVN	RFNEI	ADGT	VATV	GHLTV	APGCGG	NOVP	GEVD	FTLDL	RSRPHESLSRV 300
Qy	301	LIDRIS	VMVGE	VASQ	AGV	AAVD	VDPE	FNLS	PVQLA	PTWMDA	VREAASALQ	THRDISSGAG 360
Db	301	LIDRIS	VMVGE	VASQ	AGV	AAVD	VDPE	FNLS	PVQLA	PTWMDA	VREAASALQ	THRDISSGAG 360
Qy	361	HDSMFI	QAQV	TDVGM	VFPV	PSRAGR	SHV	PEW	TDFD	DLRKG	TEGV	VLNVKALDR 412
Db	361	HDSMFI	QAQV	TDVGM	VFPV	PSRAGR	SHV	PEW	TDFD	DLRKG	TEGV	VLNVKALDR 412

RESULT 2

US-09-950-772-6
; Sequence 6, Application US/09950772
; Patent No. US20020102713A1
; GENERAL INFORMATION:
; APPLICANT: SUZUKI, Shunichi
; APPLICANT: ONISHI, No. US20020102713A1Imasa
; APPLICANT: YOKOZAKI, Kenzo
; TITLE OF INVENTION: 5-SUBSTITUTED HYDANTOIN RACEMASE, DNA CODING FOR THE RACEMASE, AND
; TITLE OF INVENTION: PROCESSES FOR PRODUCING OPTICALLY ACTIVE AMINO ACIDS
; FILE REFERENCE: 212318USO
; CURRENT APPLICATION NUMBER: US/09/950,772
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: JP2000-278571
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: JP2001-65815
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Microbacterium liquefaciens
US-09-950-772-6

Query Match 83.4%; Score 1709; DB 9; Length 412;
Best Local Similarity 81.1%; Pred. No. 1.9e-153;
Matches 334; Conservative 38; Mismatches 40; Indels 0; Gaps 0

[illegible]

RESULT 4

US-10-425-114-49502
; Sequence 49502, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 18:22:06 ; Search time 48 Seconds
(without alignments)
2379.211 Million cell updates/sec

Title: US-10-045-063-2

Perfect score: 2049

Sequence: 1 MTLQKAQAARIKEIRELSR.....FDDLKRGTEWLRVMKALDR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep.*
- 17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	2049	100.0	412	14	US-10-334-990-7
2	1709	83.4	412	9	US-09-950-772-6
3	1709	83.4	412	14	US-10-289-360-4
4	628	30.6	472	12	US-10-425-114-49502
5	625	30.5	460	12	US-10-424-599-229959
6	526.5	25.7	483	12	US-10-424-599-158434
7	522.5	25.5	511	12	US-10-425-114-43946
8	504	24.6	448	12	US-10-424-599-229957
9	476.5	23.3	380	15	US-10-369-493-110895
10	474.5	23.2	380	15	US-10-369-493-11009
11	409.5	20.0	413	14	US-10-156-761-11032
12	143.5	7.0	319	12	US-10-282-122A-47606
13	138.5	6.8	418	12	US-10-282-122A-50358
14	124	6.1	876	12	US-10-282-122A-62122
15	123.5	6.0	1574	14	US-10-205-032-16

16	123	6.0	394	12	US-10-282-122A-62369	Sequence 62369, A
17	123	6.0	394	12	US-10-282-122A-64859	Sequence 64859, A
18	121.5	5.9	4928	16	US-10-329-148A-5	Sequence 5, Appli
19	121	5.9	875	12	US-10-282-122A-62829	Sequence 62829, A
20	121	5.9	875	12	US-10-282-122A-64519	Sequence 64519, A
21	121	5.9	875	14	US-10-080-170-464	Sequence 464, App
22	120.5	5.9	470	15	US-10-369-493-17350	Sequence 17350, A
23	120.5	5.9	713	12	US-10-282-122A-47769	Sequence 47769, A
24	120.5	5.9	5245	14	US-10-329-079-45	Sequence 45, Appl
25	119	5.8	441	14	US-10-156-761-14162	Sequence 14162, A
26	119	5.8	578	14	US-10-156-761-14026	Sequence 14026, A
27	119	5.8	714	12	US-10-282-122A-49590	Sequence 49590, A
28	118.5	5.8	393	14	US-10-156-761-12431	Sequence 12431, A
29	118.5	5.8	625	12	US-10-282-122A-63694	Sequence 63694, A
30	118	5.8	6973	12	US-09-942-035-15	Sequence 15, Appl
31	117.5	5.7	649	12	US-10-282-122A-62146	Sequence 62146, A
32	116.5	5.7	342	12	US-10-424-599-234064	Sequence 234064, A
33	116	5.7	412	12	US-10-362-327-4	Sequence 4, Appli
34	116	5.7	1085	12	US-10-335-977-6909	Sequence 6909, Ap
35	113.5	5.5	342	15	US-10-369-493-9030	Sequence 9030, Ap
36	113.5	5.5	440	14	US-10-156-761-9438	Sequence 9438, Ap
37	113	5.5	730	12	US-10-282-122A-72479	Sequence 72479, A
38	112.5	5.5	871	15	US-10-369-493-18539	Sequence 18539, A
39	112	5.5	445	14	US-10-156-761-9244	Sequence 9244, Ap
40	112	5.5	541	12	US-10-282-122A-52759	Sequence 52759, A
41	111.5	5.4	763	12	US-10-282-122A-50508	Sequence 50508, A
42	111	5.4	415	9	US-09-898-461-7	Sequence 7, Appli
43	111	5.4	415	15	US-10-275-580-2	Sequence 2, Appli
44	111	5.4	502	14	US-10-156-761-13943	Sequence 13943, A
45	110.5	5.4	362	15	US-10-369-493-7208	Sequence 7208, Ap

ALIGNMENTS

RESULT 1

US-10-334-990-7
; Sequence 7, Application US/10334990
; Publication No. US20030175910A1
; GENERAL INFORMATION:
; APPLICANT: ALTENBUCHNER, JOSEF
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: MATTES, RALF
; APPLICANT: SYLDATK, CHRISTOPH
; APPLICANT: TISCHER, WILHELM
; APPLICANT: WIESE, ANJA
; APPLICANT: WILMS, BURKARD
; TITLE OF INVENTION: WHOLE CELL CATALYST
; FILE REFERENCE: 9350-0142-0
; CURRENT APPLICATION NUMBER: US/10/334,990
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US/09/407,062
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Arthrobacter aureus
US-10-334-990-7

Query Match 100.0%; Score 2049; DB 14; Length 412;

Best Local Similarity 100.0%; Pred. No. 1e-185; Mismatches 0; Indels 0; Gaps 0;

Matches 412; Conservative 0;

QY 1 MTLQKAQAARIKEIRELSRFAEGVTRLYTTPHEAARETLIAAMKAALSVDAL 60

Db 1 MTLQKAQAARIKEIRELSRFAEGVTRLYTTPHEAARETLIAAMKAALSVDAL 60

QY 61 GNIIGREGTDELPAIVAGSHFDSVRNGMGFDGTAGVVCALAAARVLENGVYVNRHPE 120

Db 61 GNIIGREGTDELPAIVAGSHFDSVRNGMGFDGTAGVVCALAAARVLENGVYVNRHPE 120

This Page Blank (uspto)

Best Local Similarity 22.5%; Pred. No. 0.047;
Matches 106; Conservative 42; Mismatches 169; Indels 155; Gaps 21;

Qy	27	GVTRLTTYPE-----HAAARETLIAAMKA-----AALSVPREDALGNIGRREG--T 70
Db	1127	GVDRFGVHPALLDAVLHAAAAETSVVQSEARVPFSPWGRGVELRATESA---VVRARLSLTS 1183
Qy	71	DPELPAIAGVSHFDSVRNGMFDGTAGVVCALAAARVYMLNGYVNRHPPEF----- 121
Db	1184	DDELSLVAVDP-----AGRFATVDSLVTRPISRQOVRSGAIGDCLFEVEWHRKALLG 1236
Qy	122	-----IAIVEEGARFSSGMLGGRATAGLVADRELDLSLVED-----GVSVRQAATAFGL 171
Db	1237	TTAGDDLAIIV-GDGPSWPESV---RATARFATLDEPRAAVDSVPAPGSLVAAMSAAEV 1292
Qy	172	KPGELQAAAR-SAADLPAFIELHIEQGPILIEQOEI-----GVVTSIVGV----- 216
Db	1293	EGGSLPSRAQESTSDLLALVQSLADERFAESQLVVVTRAASVADSDDVADLVGASSWG 1352
Qy	217	-----RALRVAVKGRSD-----HAGTTTPMHLROD-ALVP--AALMVREYN 253
Db	1353	LLSSAQSENPGRFVLVDVDTGTPESWQALPAAVRAGEPQLALRRGVALVPRLARLTVREEG 1412
Qy	254	RFVNEIADGTVATVG-----HLTVAPGGGNQVPGEVDFTLDLRSP 293
Db	1413	SSPOLDTGTVLITGGTGALGGVVARHLVEEHGIRRLVLAGRGNAPG-----V 1462
Qy	294	HEESLRVLID---RISVMGGEVASQAGVAADVDEFFNLSP-----VOLAPTMVDVAVR 342
Db	1463	HE-----LVDELARAGAVVEVWACDVADRTDLEHLAAIPVDWFLRGIVHTAGVLADGVI 1517
Qy	343	EAASALQFTHRDISSGAGHDSMFIAQVTDYGMVFPVPSRAGRSHVPEWTDPD 394
Db	1518	GSLSA-----ADVGTVPAPKVTGAWHLHELTRDLD 1547

Search completed: May 3, 2004, 18:23:05
Job time : 24 secs

FILE REFERENCE: GTC99-03PA
CURRENT APPLICATION NUMBER: US/09/328,352
CURRENT FILING DATE: 1999-06-04
NUMBER OF SEQ ID NOS: 8252
SEQ ID NO 5309
LENGTH: 406
TYPE: PRT
ORGANISM: Acinetobacter baumannii
US-09-328-352-5309

Query Match
Best Local Similarity 6.5%; Score 133; DB 4; Length 406;
Matches 88; Conservative 67; Mismatches 170; Indels 94; Gaps 19;

QY 26 PGVTRLTYPPEHAARETLIA-AMKAAALSVREDALGNI--TCREGTDPELPAIAVGS 82
DB 46 PSVTPIDHT-----CQIMADRLAKVGHIEPMRFGVDNLWARRGT--EGPVCFAGH 97
QY 83 FD-----SVRNGMFD-GTAGVVCALBAARVMLENGVYVNRHP-----F 119
DB 98 TDVPTGRLDWNSDPPFAPEIRDGKLYGRGSADMTAL--AMVVASERFVAKHPNHKGS 156
QY 120 EFIAIVEEGARFSSGMLGGRAIGLVADRELDLSVDEGVSVRQAATAFGKLPQELQAA 179
DB 157 AFLTSDDEGP-----AVNGTVKVIETLEKRNKIKTWCLVGEPSSTHKLGDIVKN 206
QY 180 ARSAADLRAFIELHQEPILQEQIEIGVVTSSIVGVRLRVAVKGRSDHAGTTPMHLRQ 239
DB 207 GRRGS-LNAVVKVQKQGHV-----AYPHLARNIHEAS 239
QY 240 DALVPAALMVRENVNRVNIADGTVATVGHVLTVPAGGNQVPGVEVDFTLDLSPHEESLR 299
DB 240 PALAELCQTVMDNG--NEYFPATSFQISNIHAGTGATNVPGLAEVTFNFR-----YSTE 292
QY 300 VLIDRISVMVGSVAGVAGVADVDFFNLSPVQ-LAPT--MVDVAREASALQFTHRDIS 356
DB 293 VTAQLKQKVHLEIKHGLQYEI--VMNLGLPFLTPVGLYNAAQATLNTVTGETELS 350
QY 357 -SGAGHDSMFI----AQVTDVGMFVPSRAGSHVPEETDFFDLRKGFVVLVRVMKAL 410
DB 351 TSGGTSDDGFIAPTGAQVLELGVL-----NATIHQINHVHDLDPLTDIYEQILENL 404

RESULT 14
US-09-252-991A-19662
Sequence 19662, Application US/09252991A
Patent No. 6551795
GENERAL INFORMATION:
APPLICANT: Marc J. Rubenfield et al.
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
FILE REFERENCE: 107196.136
CURRENT APPLICATION NUMBER: US/09/252,991A
CURRENT FILING DATE: 1999-02-18
PRIOR APPLICATION NUMBER: US 60/074,788
PRIOR FILING DATE: 1998-02-18
PRIOR APPLICATION NUMBER: US 60/094,190
PRIOR FILING DATE: 1998-07-27
NUMBER OF SEQ ID NOS: 33142
SEQ ID NO 19662
LENGTH: 1002
TYPE: PRT
ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-19662

Query Match
Best Local Similarity 6.2%; Score 127.5; DB 4; Length 1002;
Matches 101; Conservative 44; Mismatches 125; Indels 147; Gaps 23;
QY 3 LOKAQAAARKEIRELSRFSAEQ-PGVTRLTYPPEHAAR-----E 42
DB 460 LRAAQVAVR-----LEQPRVAGLEQHEGLAPQVLGLDLDVLQAVAGQGLV 508

QY 43 TLIAAMKAAALSVREDALGNIIGRRGTDPPELPAIA-----VGSFSDSVRNGGMFDG 94
DB 509 VLVALLEGLAVEVVQ--VGD-VGRRE-----EGPAGVVEDALHQQVDDVGVGVHVVGTTTV 561
QY 95 TAGVVCALBAARVMLENGVYVNRHPEFTAIIVEEGARFSSGMLGGRAIAGLVADRELDLS 154
DB 562 VTGVLQALE-----EFLD-VQVPGQLQVGA---DRLA-----LAAL 593
QY 155 VDED-GV--SVRQAATAFGLKPGELQAAARGAADLRAFIELHQEPILQEQIEIGVVT 211
DB 594 VDRGVGVDFHQQEGHDALGLAVGALDVGAQRA-----HGPVVAQA----- 634
QY 212 SIVGVRLRVAVKGRSDHAGTTPMHLRQDALVPALMVRENVNRVNIADGTVATVGHILT 271
DB 635 -----AGEFRQH-----GVVVDGAVDARQVVRHGGQVAAAGELRTQAGV 673
QY 272 VAPGGGNQVPG-----EVDEF-----LDLRS---PHEESLRVLIDRISVMVGSVAGV 316
DB 674 EGGGTHVVEGRQOVVELOGTCLAVLLDQQAQHAHEEDLR-QFEADAVLVDEVAVVQ 732
QY 317 GVAADVDEFFNLSPVQLAPTVMVDVAREASALQFTHRDISGAGHDSMFIQAQVTDVG 373
DB 733 GLQAEVSEL-----LVALVIDGL---AQFLQVEFLQL-----GIDQLELGAFLDVG 775

RESULT 15
US-09-036-987A-5
Sequence 5, Application US/09036987A
Patent No. 6143526
GENERAL INFORMATION:
APPLICANT: Baltz, Richard H.
APPLICANT: Broughton, Mary C.
APPLICANT: Crawford, Kathryn P.
APPLICANT: Madduri, Krishnamurthy
APPLICANT: Merlo, Donald J.
APPLICANT: Treadway, Fatti J.
APPLICANT: Turner, Jan R.
APPLICANT: Waldron, Clive
TITLE OF INVENTION: Biosynthetic Genes For Spinosyn Insecticide
TITLE OF INVENTION: Production
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dow Agrosciences LLC Patent Department
STREET: 9330 Zionsville Road
CITY: Indianapolis
STATE: Indiana
COUNTRY: USA
ZIP: 46268
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/036,987A
FILING DATE: 09-MAR-1998
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Stuart, Donald R.
REGISTRATION NUMBER: 28,479
REFERENCE/DOCKET NUMBER: 50,608
TELEPHONE: (317)337-4816
TELEFAX: (317)337-4847
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 4928 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-036-987A-5

Query Match
5.9%; Score 121.5; DB 3; Length 4928;

Db 62 EQLGNNRYLVHKDATNPQ---IFTAAHLDTVPKGTVAERPFSIDGSRAYGPGVIDMKAS 118
QY 98 VVCALEARVMLENGYVNRHPFFIAIVEEE-GARFSSGMLGGRAIAGLVADRELDLSVD 156
Db 119 HVLTYAINALKOSGDLUSYQNVEITLLNCDBEIGSKTSRHLEIKYA----- 163
QY 157 EDGVSVRQAATAFGLKPGELQQAARSAAADLRAFIELHIEQGPILIEQOEIGVVVTSIVGV 216
Db 164 -----KNKAYALIMEPARANGA-----IVSARRGV 198
QY 217 RALRVAVKGRSDHAGTTPMHLRODALVPAALMVRENVRFVNEI-----ADGTVATVGH 270
Db 189 GTVELLIEGKASHGIAP-----EAGISA---IQELSYKIQALHALSRHDEGLSVNVGLI 240
QY 271 TVAPGGGNQVPGVDFTLDRSPH---EESLRVLID-----RISVMGEVASO---AGVAA 320
Db 241 S-----GGTSV-----NTVAPHARAEIDVRISTDSQGIEDIKQREVCSQVPLAGVK 288
QY 321 DVDEFFNLSPVQLAP---TWDAVAREASALQFTHRDISSGAGHDSMFIAQVTDVGMVFP 378
Db 289 TLKGINRPMWKTPESEANLIEIKQQAADLIGLAIEDISTGGSDASFTAGVGTSPVDGL 348
QY 378 PSRAGSHVPEEWTDFDLRKGTVEVLRVVKALDR 412
Db 349 GPIGGYQHSKEYLDLPLSLTERTVLFANVIKRLSQ 383

RESULT 11
US-09-328-352-6356
; Sequence 6356, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: GTC99-03PA
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 6356
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-6356

Query Match 6.9%; Score 140.5; DB 4; Length 447;
Best Local Similarity 20.8%; Pred. No. 1.4e-05;
Matches 93; Conservative 74; Mismatches 174; Indels 107; Gaps 21;
QY 12 EKEIRELSRFSASGPGVTRLTYYTPEHAAARETLIAAMKAAALSVRED-ALGNLIIGREGT 70
Db 34 ENQVIOQRHIEHPYELGNMEFKTSALVQKE-----LKSYGIQVKTGYAKTGVIILKGN 88
QY 71 DPFLPAIAGVSHFDSV-----RNGGMFDG-----TAGVWCALBAARV 107
Db 89 NPG-PIIALRADNDALPMEKSGVPFASKQKAIYQKETYVMACGDAHTAMLLGAAXI 147
QY 108 MLENGYVNRHPFFIAIVEEGA-----RFSFG-MLGGRAIAGLVADRELDLSVDED--GV 160
Db 148 LAANKDKISGTVVVFQPAPEGADIDNFTGQIGSRK---MIADGAFKDYKPEAIFGM 204
QY 161 SVRQAATAFGLKPGELQQAARSAAADLRAFIELHIEQGPILIEQOEIGVVVTSIVGVRLR 220
Db 205 HVMS-----GMKSG-----HLYYKDGAILNSAD-----HLR 230
QY 221 VAVKGRSDHAGTTPMHLRODALVPAALMVRENVRFVNEIADGT-----VATVGHILTAPGG 276
Db 231 IQVNGKQVH-GSTPW-LGRDPIYASQAMINNLQSLISRRTDLTQGMGVISIGNI---QGG 285
QY 277 --GNQVPGVDFTLDRSPHEESLRVLIDRISVMGEVASOQAAADVDEFFNLSPVQ-- 332
Db 286 TAGNVIPEQVNMIGTIRSNNEQIRENLIKSLPALVEHNAQNEVTAKV-EIAPYAPVTWN 344

QY 333 -----LAPTMVDVAREASALQFTHRDISSGAGHDSMFIAQVTDVGMVFP----- 378
Db 345 NKALTQIOLPTLAKTVGDS-----NLHVLHDNASASEDFAYYGKLMSPFFVLGATPENQD 400
QY 379 -SRAGSHVPEEWTDFDLRKGTVEVLR 405
Db 401 LSQAAPNHNPSFIVDDKALKTKGTGELHR 428
RESULT 12
US-09-252-991A-21048
; Sequence 21048, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: 107196.136
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 21048
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-21048

Query Match 6.5%; Score 133.5; DB 4; Length 434;
Best Local Similarity 20.5%; Pred. No. 6.9e-05;
Matches 84; Conservative 53; Mismatches 154; Indels 119; Gaps 14;
QY 21 FSAEGGPGVTRLTYYT-----PEHAAARETLIA-----AMKAAALSVREDALG--- 61
Db 45 FAADTPAAELLRQAEAPAYLDTLFQLVAVDSGTQAEGLGQLSALLAERLQALQAQVR 104
QY 62 -----NIIGRREGTDPFLPAIAGVSHFDSV-----RNGGMFD 93
Db 105 SAPATPSAGDNLVATLDGTGSKRFLMI--HYDTVFAAGSAAKRPPREDAAERAYGPGVAD 162
QY 94 GTAGVVCALEARVMLENGYVNRHPFFIAIVEEGARFSSGMLGGRAIAGLVADRELD 153
Db 163 AKGVAMVHLALURLQQGFRDYGRITVLFNPDEE-----TGSAGSKQLIAELA----- 211
QY 154 LVDEDCGVSVRQAATAFGLKPGELQQAARSAAADLRAFIELHIEQGPILIEQOEIGVVVTSI 213
Db 212 -----RQDYYVSEYEPDORDA-----VTVAT 232
QY 214 VGVRLRVAVKGRSDHAGTTPMHLRODALVPAALMVRENVRFVNEIADGTVATVGHLTVA 273
Db 233 NGIDGILLLEVKGRSSHAGSAPEQGRNAILLESHQLLR-----LKLGDPAKGTTLNWTLA 287
QY 274 PGG--GNQVPGVDFTLDLR-SPHEESLRVLIDRISVMGEVASOQAAADVDEFFNLSP 330
Db 288 RGGEKRNIIIPAEASAEADWRYSDPAESERVLDARKLTGERLVADTEVSLRLDKG----- 342
QY 331 VOLAPTMVDVAREASALQFTHRDISSGAGHDSMFIAQ--VTDVGMVFP 378
Db 343 ---RPLVK--NPASQRLAETAQTLYGRVGRKRIEPIAMRFGTDAAGYAYVP 387

RESULT 13
US-09-328-352-5309
; Sequence 5309, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE REFERENCE: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 18:19:20 ; Search time 22 Seconds
(without alignments)
966.814 Million cell updates/sec

Title: US-10-045-063-2
Perfect score: 2049
Sequence: 1 MTUQQAARIKEIRELSR.....FDDLKRGTEWLVKMKALDR 412

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues
Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*
1: /cgn2_6/ptodata/2/iaa/5A COMB.pep.*
2: /cgn2_6/ptodata/2/iaa/5B COMB.pep.*
3: /cgn2_6/ptodata/2/iaa/6A COMB.pep.*
4: /cgn2_6/ptodata/2/iaa/6B COMB.pep.*
5: /cgn2_6/ptodata/2/iaa/PTCUS COMB.pep.*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2049	100.0	412	4	US-09-285-055-2
2	640.5	31.3	459	4	US-09-489-039A-9613
3	639.5	31.2	495	4	US-09-252-991A-23393
4	606	29.6	425	4	US-09-134-000C-5976
5	602.5	29.4	412	4	US-09-554-998A-2
6	578.5	28.2	429	4	US-09-489-039A-12025
7	428	20.9	307	4	US-09-489-039A-12144
8	163.5	8.0	134	4	US-09-328-352-7315
9	143.5	7.0	465	4	US-09-543-681A-5198
10	141.5	6.9	383	4	US-09-328-352-7315
11	140.5	6.9	447	4	US-09-543-681A-7526
12	133.5	6.5	434	4	US-09-328-352-7315
13	133	6.5	406	4	US-09-252-991A-21048
14	127.5	6.2	1002	4	US-09-328-352-5309
15	121.5	5.9	4928	3	US-09-252-991A-19662
16	121.5	5.9	4928	3	US-09-036-987A-5
17	121.5	5.9	4928	3	US-09-370-700-5
18	120	5.9	752	4	US-09-603-207-5
19	119	5.8	496	4	US-09-252-991A-21724
20	119	5.8	478	4	US-09-252-991A-17357
21	117.5	5.7	573	4	US-09-252-991A-30721
22	117	5.7	879	4	US-09-252-991A-31661
23	116.5	5.7	879	4	US-09-252-991A-28118
24	115	5.6	573	4	US-09-252-991A-20876
25	115	5.6	1151	4	US-09-252-991A-21328
26	114.5	5.6	471	4	US-09-252-991A-27633
27	113.5	5.5	820	4	US-09-252-991A-32001

28 113 5.5 348 3 US-09-202-832-1 Sequence 1, Appli
29 113 5.5 348 3 US-09-202-832-10 Sequence 10, Appl
30 112.5 5.5 1114 4 US-09-252-991A-24965 Sequence 24965, A
31 112.5 5.5 3562 4 US-09-679-279-14 Sequence 14, Appl
32 112 5.5 433 4 US-09-252-991A-21029 Sequence 21029, A
33 112 5.5 604 4 US-09-252-991A-31648 Sequence 31648, A
34 112 5.5 926 4 US-09-252-991A-31053 Sequence 31053, A
35 111 5.4 403 3 US-08-776-246-4 Sequence 4, Appli
36 111 5.4 415 4 US-09-898-461-7 Sequence 7, Appli
37 111 5.4 643 4 US-09-252-991A-21569 Sequence 21569, A
38 111 5.4 647 4 US-09-423-439-60 Sequence 60, Appl
39 110.5 5.4 524 4 US-09-252-991A-17710 Sequence 17710, A
40 110 5.4 412 4 US-09-423-439-10 Sequence 10, Appl
41 110 5.4 490 4 US-09-328-352-4607 Sequence 4607, Ap
42 110 5.4 642 4 US-09-423-439-26 Sequence 26, Appl
43 110 5.4 643 4 US-09-423-439-16 Sequence 16, Appl
44 110 5.4 666 4 US-09-423-439-51 Sequence 51, Appl
45 110 5.4 673 4 US-09-423-439-32 Sequence 32, Appl

ALIGNMENTS

RESULT 1
US-09-285-055-2
; Sequence 2, Application US/09285055
; Patent No. 6352848
; GENERAL INFORMATION:
; APPLICANT: ALTENBUCHNER, JOSEF
; APPLICANT: MATTES, RALF
; APPLICANT: PIETZSCH, MARKUS
; APPLICANT: SYLDATK, CHRISTOPH
; APPLICANT: WIESE, ANJA
; APPLICANT: WILMS, BURKARD
; TITLE OF INVENTION: RECOMBINANT L-N-CARBAMOYLASE FROM ARTHROBACTER
; TITLE OF INVENTION: AURESCENS AND METHOD OF PRODUCING L-AMINO ACIDS
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: RECOMBINANT L-N-CARBAMOYLASE
; CURRENT APPLICATION NUMBER: US/09/285,055
; CURRENT FILING DATE: 1999-04-02
; EARLIER FILING DATE: DE 198 14 813.5
; EARLIER FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 412
; TYPE: PRT
; ORGANISM: Arthrobacter aurescens
US-09-285-055-2

Query Match 100.0%; Score 2049; DB 4; Length 412;
Best Local Similarity 100.0%; Pred. No. 2.5e-197;
Matches 412; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MTUQQAARIKEIRELSRFSAGPGVTRLVTPPEHAAARETLIAAMKAAALSVDAL 60
Db 1 MTUQQAARIKEIRELSRFSAGPGVTRLVTPPEHAAARETLIAAMKAAALSVDAL 60
QY 61 GNIIGRREGTDPPELPAIAVGSFHDSVRNGGMDFGTAGVVCALAEAAARVMLENGYVNRHPE 120
Db 61 GNIIGRREGTDPPELPAIAVGSFHDSVRNGGMDFGTAGVVCALAEAAARVMLENGYVNRHPE 120
QY 121 FTAIVVEEGARFSSGMLGGRAIAGLVADRELSDVDEGVSVRQAATATGKKEGELQAAA 180
Db 121 FTAIVVEEGARFSSGMLGGRAIAGLVADRELSDVDEGVSVRQAATATGKKEGELQAAA 180
QY 181 RAAADURAFIEHIEQPIEQEIBIGVVTSTVGVRLRVAVKGRSDHAGTTPMHLROD 240
Db 181 RAAADURAFIEHIEQPIEQEIBIGVVTSTVGVRLRVAVKGRSDHAGTTPMHLROD 240
QY 241 ALVPAALMVRENVRFVNEIADGTATVGHILTAPAGGNGVPGGEVDFTLRLSPHESLRV 300
Db 241 ALVPAALMVRENVRFVNEIADGTATVGHILTAPAGGNGVPGGEVDFTLRLSPHESLRV 300

PF 30-MAR-2001; 2001WO-US008631.
XX
XX 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX WPI; 2001-639362/73.
DR N-PSDB; AAS89217.
XX
XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
XX Claim 20; SEQ ID NO 5389; 103pp; English.
XX
XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activities. The
CC polypeptide and polynucleotide sequences have application in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG00010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 502 AA;
SQ

Query Match 7.1%; Score 146; DB 4; Length 502;
Best Local Similarity 33.0%; Pred. No. 9.2e-05;
Matches 36; Conservative 16; Mismatches 57; Indels 0; Gaps 0;

QY 48 MKAALSVREDALGNIGRGTDPDLPALAVGSHFDSVRNGMFDGTAGVVCALAEARV 107
DB 1 MAASGLETRFDEVGNLYGRNGTETYPQEVVLSGSHDITVYVNGNLDGQFGALAAWLADW 60

QY 108 MLENGVNRHPFEFTAIVEEGARSSGMLGGRAIAGLVADRELDLVD 156
DB 61 LKTQYCAPLRTVEVVTMAEEGSRFPYVFWGSKNIFGLANPDDVRNICD 109

Search completed: May 3, 2004, 18:20:31
Job time : 62 secs

CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 554 AA;

Query Match 17.1%; Score 349.5; DB 4; Length 554;
Best Local Similarity 25.1%; Pred. No. 5.7e-24;
Matches 115; Conservative 70; Mismatches 185; Indels 89; Gaps 13;
QY 9 ARIEKEIRE-----LSRFSAEKPGVRLT-----YTPHAAA---RETLIAA 47
Db 111 ARIPOEVKEKYVLKSNAPYLEILDKRVNKGTVKSLADVLGKPEIWAIGQENDIAM 170
QY 48 MKAALSVREDALGNIGRREGTDPPELPAIAGVSHFDSVRNGMFGDTA-----96
Db 171 IEYAGVG-----AMDNAIPSVKEVANF--VTKSNLEDCGVAPEKPKRSA 215
QY 97 -GVVCALEARV--MLENGVYVNRHPFEFTAIVEEGARFSSGMLGRAIAGLVADRELD 152
Db 216 SGTALWLAFLNIDIRRFQGVLFNKRTRAFNGVAHQGEQ----LVGSNRFNV---KNVN 268
QY 153 SLVDE--DQVSV-----RQAATAFGKPKGEL-----QAAARSA 183
Db 269 SVITQVSSGVLYADTKGTFNRRILVPIDISDELQTVLISHVEERAKIDDAEVHFL 328
QY 184 ADLRAFELHIEQPILEQBIQVGVTSIGVRLRVAVKGRSDHAGTTPMHLRQDALV 243
Db 329 TDIAKFVFLHIEQGVLESQSIGVNAIVGQRRYTVTLNGESNHAGITPMGVRDVTY 388
QY 244 PAALMVRVNRVNEIADGTATVATVGHLTVAPEGGNQVPGVEVDTLDIRSPHEESLRVID 303
Db 389 AFSRICHQSVKAKRMGDPVLVTFGKVEPRPNTVNVVPGKTTTIDCRHTDAAVLRDFTQ 448
QY 304 RISVMVGEVASOAGVAADVDEFNLSPVOLAPTMVDVAREASALQFTHRDISSGAGHDS 363
Db 449 QLENDRAICDEMIDIGIDLMWDEEPVPMNKELVATLTLCERKLNTRVMSHSGAGHDA 508
QY 364 MFTAQVTDVMVFPVPSRAGRSHVPEEWTDFDILRKGTVEV 402
Db 509 QIFAPRVPTCMIFIPISINGISHNPAERTNITDL--GTEI 545

RESULT 14
ABG08894
ID ABG08894 standard; protein; 435 AA.

XX
AC ABG08894;

XX
DT 13-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #8885.

XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX
OS Homo sapiens.

XX
PN WO200175067-A2.

XX
PD 11-OCT-2001.

XX
PF 30-MAR-2001; 2001WO-US008631.

PR 31-MAR-2000; 2000US-00540217.
PR 23-AUG-2000; 2000US-00649167.
XX
PA (HYSE-) HYSEQ INC.
PI Drmanac RT, Liu C, Tang YT;
XX
XX WPI; 2001-639362/73.
DR N-PSDB; AAS73081.
DR
XX
PT New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.
XX
PS Claim 20; SEQ ID NO 39253; 103pp; English.

XX
CC The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food
CC supplement. (II) and its binding partners are useful in medical imaging
CC of sites expressing (II). (I) and (II) are useful for treating disorders
CC involving aberrant protein expression or biological activity. The
CC polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human diagnostic
CC amino acid sequences of the invention. Note: The sequence data for this
CC patent did not appear in the printed specification, but was obtained in
CC electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 435 AA;

Query Match 7.7%; Score 158; DB 4; Length 435;

Best Local Similarity 33.9%; Pred. No. 5.5e-06;

Matches 39; Conservative 17; Mismatches 59; Indels 0; Gaps 0;

QY 48 MKAALSVREDALGNIGRREGTDPPELPAIAGVSHFDSVRNGMFGDTAGVVCALAAARV 107

Db 12 MAASGLETRFDEVGNLGRLNGTEYPOEVVLSGSHDITVYVNGNLGDFGALAAWLAIW 71

QY 108 MLENGVYVNRHPFEFTAIVEEGARFSSGMLGRAIAGLVADRELDLSVDEGVSV 162

Db 72 LKQYCAPLRTVEVVMAEEGSRFPYVFWGSKNIFGLANPDDVRNICDAKGNIS 126

RESULT 15

ABG25030

ID ABG25030 standard; protein; 502 AA.

XX
AC ABG25030;

XX
DT 18-FEB-2002 (first entry)

XX
DE Novel human diagnostic protein #25021.

XX
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX
OS Homo sapiens.

XX
PN WO200175067-A2.

XX
PD 11-OCT-2001.

Db 127 EGARFGMPCLSRGLSAGKLSKAEATLVARDCQSLPDAWREAGLNP-DLMGPDVLLRAS 185
Qy 188 AFIELHIEQGPTLEQEQIEIGVVTISIVGRALRVAVKGRSDHAGTTPMHLRODALVPAA- 246
Db 186 CFVELHVEQGRGLADVHPVAIATVRPHGRWAEFTQGNHAGTILIPDRHDPVVPMAQ 245
Qy 247 --LMVREVRNFVNEIADGTVAIVGHLTVAPGGGNQVPGVEVFTDLRLSPHEESRLVLD 304
Db 246 TILAAREAAE-----REGCVTVGRINVEPGGTNVIASAAATMWLDCRA--EES-----QT 293
Qy 305 ISVMYGEVASQAGVAADVDEFNLSPVQLAPTWVD--AVREASALQFTHRDI-----SSGA 359
Db 294 VTKVVEDIQQASSETAKQKQ--GCGVMTRESWTDRTAFDLRLMLQILGEDIPQLSTGA 350
Qy 360 GHDSFIAQVTDVGMVFVPSRAGRSHVPEEWTDFDLRLKRGTEVLRVMKAL 410
Db 351 GHDAATLATMTPTGMLFVRNPTGASHCPAESASDADCEAGAEALQTVLEEL 401

RESULT 12

AAB97165
ID AAB97165 standard; protein; 435 AA.

XX AAB97165;

XX 10-AUG-2001 (first entry)

XX Pseudomonas NCC hydrolase enzyme.

XX Pseudomonas; 2-aminothiazoline-4-carboxylate hydrolase; ATC hydrolase;
KW enzyme; N-carbamoyl-L-cysteine hydrolase; NCC hydrolase;
KW 2-aminothiazoline-4-carboxylic acid; L-cysteine production.

XX Pseudomonas sp.

XX JP2001069985-A.

XX 21-MAR-2001.

XX 01-SEP-1999; 99JP-00247999.

XX 01-SEP-1999; 99JP-00247999.

XX (NIRI-) NIPPON RIKAGAKU YAKUIN KK.

XX WPI; 2001-347865/37.

XX N-PSDB; AAH27511.

XX A gene participating to the formation of L-cysteine from 2-
PT aminothiazoline-4-carboxylic acid and an enzyme encoded by it.

XX Claim 4; Fig 11-12; 16pp; Japanese.

XX The invention relates to a 2-aminothiazoline-4-carboxylate (ATC)
CC hydrolase gene having a fully defined 406-957 base pair sequence given in
CC the specification. The pseudomonas ATC hydrolase and N-carbamoyl-L-
CC cysteine (NCC) hydrolase enzymes are encoded by the gene. These may be
CC reacted with 2-aminothiazoline-4-carboxylic acid to produce L-cysteine.
CC The present sequence encodes the NCC hydrolase

SQ Sequence 435 AA;

Query Match 19.3%; Score 395.5; DB 4; Length 435;
Best Local Similarity 30.8%; Pred. No. 1.8e-28;
Matches 136; Conservative 52; Mismatches 174; Indels 79; Gaps 15;
Qy 2 TLQQAARIEKIELSRFSA-EGGVTRLYTPEHAARETLIAAMKAALSVRDAL 60
Db 24 TALRVNGARLWDSLERMARIGATQGGVCRALTDELREGDRLFQWAREAGCTIRVDQM 83
Qy 61 GNIIGRRETDPELPAIVAGSHFDSVRNGMFDGTAGVCALEAARVMLENGVVRHPPE 120
Db 84 GNVFARR-GRNPDAAPVLGTGSHADTQPTGGRYDGIYGLGGLVIRTLDAGVTRTEHPVE 142

Qy 121 FIATVEEGARFSSGMLGGRAIAGLVADRELDLSLVEDGVSVRQAATAFGLKPGELQAAA 180
Db 143 AVITWNEGSPARDDLVGR-LCGFVPARIALAHRHGYDDRRGADRRLRG-----RRAG 198
Qy 181 RSAADLAFIEHIEQGPILQEQIEIGVVTISIVGRALRVAVKGRSDHAGTTPMHLROD 240
Db 199 RRHA-VHAAAYELHIEQGAILERSKGTIGVVTAGQQRWYEIELVGVDAHAGTTPMDLRD 257
Qy 241 ALVPAALMVREVRNFVNEIADGTVAIV-----GHLTVAPGGGNQVPGVEVFTDLRL 291
Db 258 ALAGGA-----HDYVRRGAGAPRAAWARHGRHGRDGRGAQLA-----QHVPGRCFTVEFR 307
Qy 292 SPHEESRLVLDRLISVMYGEVASQAGVAADVDEFNLS-----PVQLAPT 336
Db 308 HP-----DAGVLAELDRNCALSWPRGAGGAARAHDRADRLLCAG 346
Qy 337 MVDVRE-----AASALQFTHRDISGAGHDSMFTAQVTDVGMVF-----VPSRAGRSHV 386
Db 347 AVRASMRHPVGAAVMLDLPDHDDIVSGAGHDACYLA-CRAHGHDFRAVRRRVESQRGAI 405
Qy 387 PEEWTFDFFDLRLKRGTEVLRVM 407
Db 406 P-EWAE-----AGANVLLQVM 420

RESULT 13

ASG28578

ID ASG28578 standard; protein; 554 AA.

XX ASG28578;

XX 18-FEB-2002 (first entry)

XX Novel human diagnostic protein #28569.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;
KW food supplement; medical imaging; diagnostic; genetic disorder.

XX Homo sapiens.

XX WO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US008631.

XX 31-MAR-2000; 2000US-00540217.

XX 23-AUG-2000; 2000US-00649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

XX N-PSDB; AAS92765.

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity.

XX Claim 20; SEQ ID NO 58937; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and polypeptide (II)
CC sequences. (I) is useful as hybridisation probes, polymerase chain
CC reaction (PCR) primers, oligomers, and for chromosome and gene mapping,
CC and in recombinant production of (II). The polynucleotides are also used
CC in diagnostics as expressed sequence tags for identifying expressed
CC genes. (I) is useful in gene therapy techniques to restore normal
CC activity of (II) or to treat disease states involving (II). (II) is
CC useful for generating antibodies against it, detecting or quantitating a
CC polypeptide in tissue, as molecular weight markers and as a food

DR N-PSDB; AAS59520.
XX
XX Propionibacterium acnes polypeptides and nucleic acids useful for
PT vaccinating against and diagnosing infections, especially useful for
PT treating acne vulgaris.
XX
XX
XX Example 1; SEQ ID NO 4651; 1069pp; English.
XX
XX Sequences AAU39105-AAU68017 represent Propionibacterium acnes immunogenic
CC polypeptides. The proteins and their associated DNA sequences are used in
CC the treatment, prevention and diagnosis of medical conditions caused by
CC P. acnes. The disorders include SAPHO syndrome (synovitis, acne,
CC pustulosis, hyperostosis and osteomyelitis), uveitis and endophthalmitis.
CC P. acnes is also involved in infections of bone, joints and the central
CC nervous system, however it is particularly involved in the inflammatory
CC lesions associated with acne vulgaris. A method for detecting the
CC presence or absence of P. acnes in a patient comprises contacting a
CC sample with a binding agent that binds to the proteins of the invention
CC and determining the amount of bound protein in the sample. The
CC polypeptides may be used as antigens in the production of antibodies
CC specific for P. acnes proteins. These antibodies can be used to
CC downregulate expression and activity of P. acnes polypeptides and
CC therefore treat P. acnes infections. The antibodies may also be used as
CC diagnostic agents for determining P. acnes presence, for example, by
CC enzyme linked immunosorbent assay (ELISA). Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
XX

Sequence 403 AA;

Query Match 21.7%; Score 444; DB 4; Length 403;
Best Local Similarity 32.8%; Pred. No. 4.1e-33;
Matches 135; Conservative 50; Mismatches 198; Indels 28; Gaps 10;

QY 8 AARIEKEIRLSRFSAEFGVTRLTTPPEHAARETLIAAMKAAALSVREDALGNIGRR 67
DB 11 AASLLAEIAAIGR--NEDGSYSRFLRPEEVALREWFVAKATELGIAIVTDANANIWAW 68
QY 68 EGTDPPELPAIAGVSHFDVSVRNGMFGDTAGVVCALAEARVMLENGVYVNRHPPEFIAIVE 127
DB 69 GTPGPD--AVLTGSHLDSVPGGGAYDGLGVSSLVAVAKMQBAGIMPAPFAVAMADE 126
QY 128 EGARFSSGMLGGRRAIAGLVADRELDSDVDGVSVRQAATAFGLKPGELQAAARSADLR 187
DB 127 EGARFGMPLGSRKLAGSKUSKAETHLVARDGQSLDPDAWREAGLN-P-DLMGDDVULLEAS 185
QY 188 AFTELHIEQGPILQEQIEIGVVTISVGVRAKRVAVKGRSDHAGTTPMHLRQDALVPAA- 246
DB 186 CFVELHVEQGRGLADGVHPVAIATAVRPHGRWRAEFTGGNHAGTTLIPDRHDPVVPMAQ 245
QY 247 --LMVREVRNFNEIADGTAVTGHUTVAPGGNQVPGVEVDFTLDRSPHERSLRVLIDR 304
DB 246 TILAAREAAE-----REGCVTTVGRINVPBGNTVIAASAATMWLDCRA--EES-----QT 293
QY 305 ISVMGEVASQAGVADVDDEFNLSVPQLAPTWD--AVREASALQFTHRID---SSCA 359
DB 294 VTKVVEDIQOASETAADKQ---GCQVWTRESWTDRTAFDLRARMLQILGEDIPQUSTGA 350
QY 360 GHDSMFIAQVTDVGMVFPVPSRAGRSHVPEWTFDLDLRKGTVEVLRVMKAL 410
DB 351 GHDAATLATMTPTGMLFVRNPTGASHCPAESASDADCEAGAEALQTVLEEL 401

RESULT 11
ABM39975
ID ABM39975 standard; protein; 403 AA.
XX
XX ABM39975;
AC
XX
XX 20-OCT-2003 (first entry)
XX
XX Propionibacterium acnes predicted ORF-encoded polypeptide #4651.

XX
XX Acne vulgaris; antiseborrhoeic; dermatological; antibacterial;
KW immunostimulant; immune response; vaccine.
XX
XX Propionibacterium acnes.
OS
XX
XX WO2003033515-A1.
XX
XX 24-APR-2003.
XX
XX 11-OCT-2002; 2002WO-US032727.
XX
XX 15-OCT-2001; 2001US-00978825.
XX
XX (CORI-) CORIXA CORP.
XX
XX Mitcham JL, Skeiky YAM, Persing DH, Bhatia A, Maisonneuve JL;
PI Zhang Y, Wang S, Jen S, Lodes WJ, Benson DR, Jones R, Carter D;
PI Barth B, Vallieue-Douglas J;
XX
XX WPI: 2003-381789/36.
DR N-PSDB; ACP64449.
XX
XX New Propionibacterium acnes polypeptides and polynucleotides encoding the
PT polypeptide, useful for diagnosing, preventing or treating acne vulgaris,
PT or for stimulating an immune response specific for a P. acnes protein.
XX
XX Example 1; SEQ ID NO 4651; 1481pp; English.

CC The invention relates to an isolated polynucleotide (ACF64435-ACF64733)
CC encoding a Propionibacterium acnes protein. The invention also relates to
CC polypeptides encoded by the polynucleotides (ABM35624-ABM64536) and to
CC immunogenic fragments of P. acnes polypeptides. The invention
CC additionally encompasses expression vectors and host cells comprising a
CC polynucleotide of the invention; antibodies against polypeptides of the
CC invention; fusion proteins comprising a polypeptide of the invention; a
CC method for stimulating an immune response specific for a P. acnes
CC polypeptide and an isolated T cell population comprising T cells prepared
CC via this method; a vaccine composition (comprising P. acnes polypeptides,
CC polynucleotides, antibodies, fusion proteins, T cell populations, or
CC antigen-presenting cells that express the polypeptide); a method and kit
CC for detecting or determining the presence or absence of P. acnes in a
CC patient; and a method for inhibiting the development of P. acnes in a
CC patient. The P. acnes polypeptides, polynucleotides, antibodies, fusion
CC proteins, T cell populations or antigen-presenting cells that express the
CC polypeptides are useful for diagnosing, preventing or treating acne
CC vulgaris, or for stimulating an immune response specific for a P. acnes
CC protein. The polynucleotides can also be used as probes or primers for
CC nucleic acid hybridisation. The vaccine composition is useful for the
CC stimulation of an immune response against P. acnes, or for treating acne,
CC and the kit is useful for performing a diagnostic assay. The present
CC sequence represents a polypeptide predicted to be encoded by an ORF (open
CC reading frame) contained within the P. acnes polynucleotides of the
CC invention. Note: The sequence data for this patent did not form part of
CC the printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
XX Sequence 403 AA;

Query Match 21.7%; Score 444; DB 6; Length 403;
Best Local Similarity 32.8%; Pred. No. 4.1e-33;
Matches 135; Conservative 50; Mismatches 198; Indels 28; Gaps 10;

QY 8 AARIEKEIRLSRFSAEFGVTRLTTPPEHAARETLIAAMKAAALSVREDALGNIGRR 67
DB 11 AASLLAEIAAIGR--NEDGSYSRFLRPEEVALREWFVAKATELGIAIVTDANANIWAW 68
QY 68 EGTDPPELPAIAGVSHFDVSVRNGMFGDTAGVVCALAEARVMLENGVYVNRHPPEFIAIVE 127
DB 69 GTPGPD--AVLTGSHLDSVPGGGAYDGLGVSSLVAVAKMQBAGIMPAPFAVAMADE 126
QY 128 EGARFSSGMLGGRRAIAGLVADRELDSDVDGVSVRQAATAFGLKPGELQAAARSADLR 187

28	Db	AFCLQNDQLDELSSFSDA	SPSVTRVLYTDDKDVSR	RVYVKNL	MALAGLTVREDAV	GNIFGK	87		
67	Qy	RECTOPELPAI	AVGSHFSDVSRNGM	FMDCTAGV	WCALEAAR	VMLENGVYVNRHP	FEFFIAIVE 126		
88	Db	WDGLEPNLP	PAVATGSHIDAI	PSYKYG	DGVGVGLGAIE	AINVLKESGFKP	KRSLEILFTS 147		
127	Qy	EGEARPSSG	MLGGAIAGL--	VARELDSL	VDEGVS	VROQAATAFGL---	KPELQQAAR 181		
148	Db	EBPFRFGIS	CLGSRLLAGS	KELAEAL	KTVTVVDG	QNVSFIEAARS	SAGVAEDKDDLSVFL 207		
182	Qy	SAADLRA	FAELHIEG	PITILEQE	IEIGVVT	SI VGV	RALRVAVKGRSDHAGTTPMHLRQDA 241		
208	Db	KKGSYFA	LELHIEG	PILEDEGL	DIGVVT	ATAAPASL	KVEFEGNGHGAGVILMPYRND 267		
242	Qy	LVPAALM	VRENVRF	NEIAD--	GTVATV	GHLTVAPGGG	NOVPGEVDTFLDLRS	PHESSLRV 300	
268	Db	GLAAELAL	AVEKHV	ESESIT	DTVTG	VI LELH	FGAINSIPSKSHLEIDTRD	IDEARRNT 327	
301	Qy	LIDRISVM	YGEVAS	QAAGVAD	VDEFN	LSPVL	APTMY--DAVREAA	SALOTHRD	ISSGA 359
328	Db	VIKKIQES	ANTIAK	RKVKLS	SEFKIV	NQDPAL	SDKLVI	KKMAEAAATELN	LSHKWMISRA 387
360	Qy	GHDSP	FIQVTD	GVGVF	PSRAG	SRHVP	EEWTD	FD	DLRKGTEVV 403
388	Db	YHDSLF	MFARIS	PMGIMF	IPC	YKGYSH	KPEEYSS	PEDMANG	VKVL 431

RESULT 9	
AAV06396	
ID	AAV06396 standard; protein; 412 AA.
XX	
XX	AAV06396;
XX	
XX	
DT	20-SEP-1999 (first entry)
XX	
DE	Bacillus subtilis metalloprotease YurH.
XX	
XX	
KW	Metalloprotease; protease; YurH; detergent; surfactant; cleaning;
KW	textile; feedstuff; animal feed; host cell.

The present sequence represents a novel metalloprotease (MP), designated YurH, of *Bacillus subtilis*. YurH DNA (see AAX59333) was identified via a BLAST search of *B. subtilis* genomic DNA. The deduced protein sequence shows identity to the MP succinyl-diaminopimelate desuccinylase from *Escherichia coli*. An expression vector including YurH DNA and a host cell comprising the vector are claimed. Also claimed are a cleaning composition, an animal feed and a composition for the treatment of a textile, all comprising YurH. Gram positive microorganisms having a mutation or deletion of all or part of YurH DNA are used as host cells for expression of a homologous or heterologous protein, such as a hormone, growth factor, cytokine or enzyme, especially a protease.

[illegible]

RESULT 10	
AAU43456	
ID	AAU43456 standard; protein; 403 AA.
XX	
XX	AAU43456;
XX	
XX	AC
XX	
XX	27-FEB-2002 (first entry)
XX	
XX	Propionibacterium acnes immunogenic protein #4352.
XX	
XX	SAPHO syndrome; synovitis; acne; pustulosis; hypertosis; osteomyelitis;
KW	uvuitis; endophthalmitis; bone; joint; central nervous system; ELISA;
KW	inflammatory lesion; acne vulgaris; enzyme linked immunosorbent assay;
KW	dermatological; osteopathic; neuroprotectant.
XX	
XX	Propionibacterium acnes.
OS	
XX	WO200181581-A2.
FN	
XX	
XX	01-NOV-2001.
PD	
PD	
XX	
XX	20-APR-2001; 2001WO-US012865.
PF	
XX	
XX	21-APR-2000; 2000US-0199047P.
PR	
PR	02-JUN-2000; 2000US-0208841P.
PR	
PR	07-JUL-2000; 2000US-0216747P.
XX	
XX	
XX	(CORI-) CORIXA CORP.
PA	
XX	
XX	Skeiky YAW, Persing DH, Mitcham JL, Wang SS, Bhatia A;
FI	
PI	L'maisonneuve J, Zhang Y, Jen S, Carter D;
XX	
XX	WPI; 2001-616774/71.
DR	

Db 297 WKAIARAETIAKERNVVTTERLQEMPPVLCSDEVKRAEAAACQKLGYPFWLPSGAH 356
QY 362 DSMFIAQVTDVGMVFPVPSRAGSHVPEWTFDDDLKRGTEVV 403
Db 357 DSVQLAPICPIGMIFVRSQDGVSHSPAEMSTKEDCAAGAEVL 398

RESULT 7
AAW03544
ID AAW03544 standard; protein; 409 AA.
AC AAW03544;
XX
DT 16-OCT-2003 (revised)
DT 04-MAR-1997 (first entry)
XX
DE Bacillus stearothermophilus amino acid amidohydrolase.
XX
KW Amino acid amidohydrolase; carbamoylase; L-methionine; stereospecific;
KW thermostable; N-carbamoyl; cam gene.
XX
OS Geobacillus stearothermophilus.
XX
FH Key Location/Qualifiers
FT Misc-difference 1 /note= "corresponds to a TTG codon"
FT
XX
PN FR2728905-A1.
XX
PD 05-JUL-1996.
XX
PF 29-DEC-1994; 94FR-00015838.
XX
PR 29-DEC-1994; 94FR-00015838.
XX
PA (RHON) RHONE POULENC NUTRITION ANIMALE.
XX
PI Dion M, Batisse N, Weigel P, Lecocq FM, Hallett JN, Sakanyan V;
XX
DR WPI; 1996-335881/34.
DR N-PSDB; AAT39170.
XX
PT New stereospecific, heat-stable amino acid amidohydrolase - from
PT B.stearothermophilus and related DNA, esp. for prodn. of L-Met.
XX
PS Claim 1; Page 9-11; 17pp; French.
XX
CC PetI fragments of Bacillus stearothermophilus NCIB 8224 genomic DNA were
CC inserted into pBR322 and used for transforming E.coli hosts. Selection
CC was on medium which contained tetracycline but lacked arginine. All
CC positive transformants carried plasmids with a 4.7 kb insert. The present
CC sequence is that of an amino acid amidohydrolase encoded by the fragment;
CC it is stereospecific and heat-stable (maximum activity at 55-60 deg.C).
CC The enzyme hydrolyses N-carbamoyl amino acid derivs. to L-amino acids and
CC is particularly useful for production of L-methionine. (Updated on 16-OCT
CC -2003 to standardise OS field)
XX
SQ Sequence 409 AA;
Query Match 35.1%; Score 719.5; DB 2; Length 409;
Best Local Similarity 38.6%; Pred. No. 3.4e-59;
Matches 155; Conservative 71; Mismatches 165; Indels 11; Gaps 4;

QY 7 QAARIEKIRELSRFSAE-GPGVTRITYTPEHAARETLIAAMKAALSVREDALGNIG 65
Db 3 QGERLWQRLMELGELGVCKPQSGGVTRLSFTAERRAKDLVASVYREAGLFVYEDAGNLIG 62
QY 66 RREGTDPELPATAVGSHFDSVRNGMFDGTAGVVCALAEARVWLENGYVNRHFFFIATV 125
Db 63 RREGTNPDAVTVLVGSHLSDVYNGGCFDGLGVLAGEVVQTMNEHGVVTHHPPIEVAFT 122
QY 126 EEEGARFSSGMLGGRAIGLADVADRELDSDVEDGVSVRQAATAFGLKPGELQAARSAAD 185

Db 123 DEEGARFRFGMIGSRAMAGTLPPEALECR-DAEGISLAEMKQAGLDPDLPOARKPGT 181
QY 186 LRAFTELHIEQPILEQEQIEIGVVTISIVGRALRVAVKGRSDHAGTTPMHLRODALVPA 245
Db 182 VKAYVELHIEQQRVLEEAAGLPVGIIVTGIAGLIHWKFTIAGPAEHAGATPMSLRDRDMAA 241
QY 246 ALMW-----REVNRVNEIADGTVATVGHILTVAPEGGNQVPGVEDFTDLRSPEEISLRLV 301
Db 242 AQIIIVIEEARR-----TGTIVGTGQLHVYPGGINVIPEVEFVLDLRLDKAEVRDQV 296
QY 302 IDRISVMGEVASQAGVAADVDEFFNLSPVLQAPTMVDAREAAASALQFTHRDISSGAGH 361
Db 297 WKAIARAETIAKERNVRLTTERLQEMADVLCSVVKQAAERACKQLGYPPFWLPSGAH 356
QY 362 DSMFIAQVTDVGMVFPVPSRAGSHVPEWTFDDDLKRGTEVV 403
Db 357 DGVQLAPICPIGMIFVRSQDGVSHSPAEMSTKEDCAVGAELV 398

RESULT 8
ABB93697
ID ABB93697 standard; protein; 441 AA.
XX
AC ABB93697;
XX
DT 31-MAY-2002 (first entry)
XX
DE Herbicidally active polypeptide SEQ ID NO 2908.
XX
KW Herbicidal; plant; agriculture; herbicide.
XX
OS Arabidopsis thaliana.
XX
PN WO200210210-A2.
XX
PD 07-FEB-2002.
XX
PF 28-AUG-2001; 2001WO-EP009892.
XX
PR 28-AUG-2001; 2001WO-EP009892.
XX
PA (FARB) BAYER AG.
XX
PI Tietjen K, Weidler M;
XX
DR WPI; 2002-269010/31.
XX
PT Identifying plant target proteins for herbicidally active compounds,
PT comprising aligning and comparing nucleic acid or amino acid sequences
PT from plant with nucleic acid or amino acid sequences from non-plant
PT organisms.
XX
PS Claim 5; SEQ ID NO 2908; 261pp + Sequence Listing; English.
XX
CC The invention relates to identifying target proteins (ABB90790-ABB94016)
CC for herbicidally active compounds, comprising aligning and comparing
CC nucleic acid or amino acid sequences from plant with nucleic acid or
CC amino acid sequences from non-plant organisms using suitable search
CC parameters, where plant sequences having an E-value greater by a factor
CC of 3 than the E-value of most similar non-plant sequences are selected.
CC The polypeptides or nucleic acids encoding them are useful for
CC identifying modulators. The identified modulators are useful as
CC herbicides
XX
SQ Sequence 441 AA;
Query Match 31.0%; Score 636; DB 5; Length 441;
Best Local Similarity 37.1%; Pred. No. 3.1e-51;
Matches 150; Conservative 71; Mismatches 175; Indels 8; Gaps 5;

QY 8 AARIEKIRELSRFS-ABGPGVTRITYTPEHAARETLIAAMKAALSVREDALGNIGR 66

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 3, 2004, 18:14:00 ; Search time 58 Seconds
(without alignments)
2007.062 Million cell updates/sec

Title: US-10-045-063-2
Perfect score: 2049
Sequence: 1 MTLQKAQAARIEKRELSR.....FDDLKRGTEWLVKALDR 412

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A Geneseqp_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2049	100.0	412	4 AAU01557	Aau01557 Arthrobac
2	1709	83.4	412	5 ABB93394	Abb93394 Amino aci
3	1709	83.4	412	5 ABB80945	Abb80945 M. liquef
4	858	41.9	423	5 ABB47375	Abb47375 Listeria
5	727	35.5	414	2 AAR15484	Aar15484 Protein e
6	726.5	35.5	409	2 AAR25693	Aar25693 Heat resi
7	719.5	35.1	409	2 AAU01544	Aau01544 Bacillus
8	636	31.0	441	5 ABB93697	Abb93697 Herbicida
9	602.5	29.4	412	2 AAU06396	Aay06396 Bacillus
10	444	21.7	403	4 AAU43456	Aau43456 Propionib
11	444	21.7	403	6 ABB39975	Abm39975 Propionib
12	395.5	19.3	435	4 AAB97165	Aab97165 Pseudomon
13	349.5	17.1	554	4 ABG28578	Abg28578 Novel hum
14	158	7.7	435	4 ABG08894	Abg08894 Novel hum
15	146	7.1	502	4 ABG25030	Abg25030 Novel hum
16	146	7.1	502	4 ABG25687	Abg25687 Novel hum
17	143.5	7.0	319	6 ABU19682	Abu19682 Protein e
18	143.5	7.0	465	6 ADA36028	Ada36028 Acinetoba
19	141	6.9	346	2 AAR67430	Aar67430 Thermosta
20	140.5	6.9	447	6 ADA35069	Ada35069 Acinetoba
21	139	6.8	349	4 AAB96717	Aab96717 Putative
22	138.5	6.8	418	6 ABU22434	Abu22434 Protein e
23	133	6.5	406	6 ADA34022	Ada34022 Acinetoba
24	130.5	6.4	415	6 ABM71060	Abm71060 Staphyloc
25	127.5	6.2	383	5 ABP65813	Abp65813 Bifidobac

26	124	6.1	401	4 ABB62639	Abb62639 Drosophil
27	124	6.1	439	5 ABR38866	Abr38866 A. niger
28	124	6.1	876	6 ABU34198	Abu34198 Protein e
29	123.5	6.0	1574	6 AAE37002	Aae37002 Micromono
30	123	6.0	394	6 ABU34445	Abu34445 Protein e
31	123	6.0	394	6 ABU36935	Abu36935 Protein e
32	121.5	5.9	4924	4 AAB70968	Aab70968 S. spinos
33	121.5	5.9	4928	2 AAY39300	Aay39300 Spnd a po
34	121	5.9	875	5 ABU05813	Abu05813 M. tuberc
35	121	5.9	875	6 ABU36595	Abu36595 Protein e
36	121	5.9	875	6 ABU34905	Abu34905 Protein e
37	120.5	5.9	713	6 ABU19845	Abu19845 Protein e
38	119	5.8	714	6 ABU21666	Abu21666 Protein e
39	118.5	5.8	351	4 AAB96205	Aab96205 Putative
40	118.5	5.8	625	6 AAB35770	Abu35770 Protein e
41	118	5.8	511	3 AAB14148	Aab14148 Bordetell
42	118	5.7	6973	7 ADC26982	Adc26982 Sorangium
43	117.5	5.7	649	6 ABU34222	Abu34222 Protein e
44	117.5	5.7	842	1 AAP93712	Aap93712 Sequence
45	116	5.7	412	5 AAU91064	Aau91064 Neisseria

ALIGNMENTS

RESULT 1
AAU01557
ID AAU01557 standard; protein; 412 AA.

XX AC

AAU01557;

18-JUL-2001 (first entry)

XX Arthrobacter aureus carbamoylase (hyc).

XX Hydantoinase; racemase; carbamoylase; whole cell catalyst; L-amino acid;
KW D-amino acid; 5-monosubstituted hydantoin; racemisation; hydrolysis;
KW bacterial cell; Escherichia coli; fermentation; purification; hyc;
KW reproduction rate.

XX Arthrobacter aureus.

XX WO200123582-A1.

XX 05-APR-2001.

XX 31-AUG-2000; 2000WO-BF008473.

XX 28-SEP-1999; 99US-00407062.

XX (DEGS) DEGUSSA-HUELS AG.

XX (UYST-) UNIV STUTTGART.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX Altenbuchner J., Mattes R., Syldatk C., Wiese A., Wilms B;

XX Bommarius A., Tischer W;

XX WPI; 2001-266169/27.

XX N-PSDB; AAS02276.

XX New whole cell catalyst for degrading hydantoin into amino acids,
comprises a hydantoinase, a racemase and a carbamoylase.

XX Example; Page 34-35; 5lpp; English.

XX The sequence represents an Arthrobacter aureus carbamoylase (hyc). A
whole cell catalyst for the conversion of 5-monosubstituted hydantoin to
L- or D-amino acids, comprises cloned genes encoding for a hydantoinase,
a racemase and a carbamoylase. These genes are overexpressed in the cell
according to their turnover rates. The bioconversion consists of chemical
and/or enzymatic racemisation of 5-substituted hydantoin, ring opening
hydrolysis achieved by a hydantoinase, and hydrolysis of the N-carbamoyl
amino acid produced by hydantoinase to the amino acid by carbamoylase.

source

```
1. .669
/organism="Pinus pinaster"
/mol_type="mRNA"
/strain="ecotype: Corsican"
/db_xref="taxon:71647"
/clone="FP006D11"
/tissue_type="differentiating xylem"
/dev_stage="adult"
/clone_lib="Pinus pinaster differentiating xylem adult"
RI: Site 2: Xho I; A composite cDNA library was made with
mRNA isolated from normal, compression, opposite, early
and late wood of Maritime pine uni-directionally cloned
into Uni-Zap XR using the ZAP-cDNA Synthesis kit
(Stratagene). pBluescript SK(-) plasmids were obtained by
in vivo mass excision. The nucleotide sequence of the
5'end was obtained by automated sequencing with the T3
primer by GENOME EXPRESS, Meylan, France"
```

ORIGIN

```
Query Match      6.1%; Score 76; DB 13; Length 669;
Best Local Similarity 50.8%; Pred. No. 3.4e-05;
Matches 181; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 82 GTTACCGGCTGACCTACACTCCAGAGCATGCCGCGCGGGAACGCTCATTTGGCGCT 141
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 142 ATGAAAGCGCGCCTTGAGCGCTTCGTGAAGACGCACTCGGAACATCATCGCCGAGCT 201
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 ATGGAGGAGCTGGTCTTCGTCCGTGAGGATGCTGTAGGAAATATATTTGGTCGTTGG 391
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 202 GAAGGCACCTGATCCGGAGCTTCCTCGATCGCGGTCGGTTACACTTCGATTCGTCCGA 261
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 392 GAGGAGCAATCCAGAGTTATCTGCTGTAGCAACTGGATCTCATATTGATGCTATTCCA 451
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 262 AACGGGGGATGTTTGTATGGCACTGCAAGCGTGGTGCGCCCTTGAGGCTGCCCGGGTG 321
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 452 TACTCAGGCAATTTGATGGAGTTGTTGGGTTCTGGGTGCCATTGAAGCTATAAATGTG 511
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 322 ATGCTGGAGAACGCTAGTGAATCGGCATCCATTTGAGTTTCATCGCATCGTGGAGGAG 381
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 512 CTGAGAAGTTGGCATTCAACCAAGAGATCACTGGAGGTGATTTATGTTACCTCTGAG 571
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 382 GAAGGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGGCCCATTTCCAGGGTTGGT 437
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 572 GAGCCTACACGCTTTGGAATTCGCTGCCCTTGAAGCCGTTTATTGGCAGGGAGTGT 627
Db ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

Search completed: May 4, 2004, 16:01:56
Job time : 3701 secs

```
Query Match      6.2%; Score 76.6; DB 12; Length 514;
Best Local Similarity 51.1%; Pred. No. 2.3e-05;
Matches 179; Conservative 0; Mismatches 171; Indels 0; Gaps 0;

QY 82 GTTACCGGTGACCTACATCCAGACATGCCCGCGGGGAAACGCTCATTTGGCGGT 141
DB 138 GTAACCAAGGCTCTGTATCTACGACAGGATGTTTGTAGTGCAGGATGTAATAAACCCAG 197

QY 142 ATGAAGCGCGCGCTTGGGCTTCGTAAGACCATCGGAACATCATCGCCGCGGT 201
DB 198 ATGAACCTTGTGCGCTATCTGTGAGAGAGATGCTGTGTGTAACATATTGGTCGTGG 257

QY 202 GAAGGCACTGATCCGAGCTTCTCGGATCGCGTTCGTTGCACACTTCGATTCTGTCCGA 261
DB 258 GATGGCCCTGAACTGAGCTTGTGCGAGTTGCAACAGGTTGCGACATGATGATATACCT 317

QY 262 AACGGCGGGATGTTTGTATGACATGTCAGGCGTGTGTGCGCCCTTGAGGCTGCCCGGGT 321
DB 318 TACTCGGGAAATATGATGAGTGTGTTGGTGTGTTTGTAGTGTCTATTGAAGCCATCAGATC 377

QY 322 ATGCTGAGAACGCTACGTGAATCGGCATCCATTTCAGTTTCATCGGCATCGTGGAGGAG 381
DB 378 CTGAAAGAGTCTGGTGTGTTTAAACCTAGAAAGACCATTTGGAAGTCATATCATTCACATCAGAA 437

QY 382 GAAGGGCGCGCTTTCAGCAGTGGCATGTTGGGGCGCGGCGCATTTGCGAGG 431
DB 438 GAACCAACACGCTTTGGAATAGGTGCTTGGGAAGCGCCCTATTGGCTGG 487

RESULT 14
BI973078      565 bp      mRNA      linear      EST 29-NOV-2001
LOCUS      s986f04.y2 Gm-cl084 Glycine max cDNA clone GENOME SYSTEMS CLONE
DEFINITION      ID: Gm-cl084-1040 5' similar to TR:085664 085664 N-CARBAMYL-L-AMINO
ACID AMIDOHYDROLASE. ; mRNA sequence.
BI973078
ACCESSION      BI973078.1 GI:16347483
VERSION
KEYWORDS
SOURCE
ORGANISM
Glycine max (soybean)
Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE
1. (bases 1 to 565)
Shoemaker,R., Keim,P., Vodkin,L., Erpelding,J., Coryell,V.,
Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
McCann,R., Waterston,R. and Wilson,R.
Public Soybean EST Project
Unpublished (1999)
Contact: Shoemaker R/Public Soybean EST Project
Public Soybean EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
This clone is available through: ResGen, Invitrogen Corp. 2130
South Memorial Parkway Huntsville, AL 35801 For further information
call: (800)-533-4363 or contact via email: ccu@resgen.com
High quality sequence stop: 424.
Location/Qualifiers
FEATURES
source
1. .565
/organism="Glycine max"
/mol_type="mRNA"
/db_xref="taxon:3847"
/clone="GENOME SYSTEMS CLONE ID: Gm-cl084-1040"
/tissue_type="Etiolated hypocotyls (Williams 82)"
/lab_host="DH10B"

/clone lib="Gm-cl084"
/notes="Vector: pBluescript II SK+; Site 1: EcoRI; Site 2:
XhoI. The cDNA library was constructed by M. Bhattacharyya
from mRNA isolated from etiolated hypocotyls from the
cultivar Williams 82. Tissue was inoculated with
Phytophthora soyae race 1 and tissues were harvested 2 and
4 hours following infection. The library is the pool of
these two time points. Complementary DNA was synthesized
from mRNA using a primer consisting of a poly(dT) sequence
with a XhoI restriction site. EcoRI adapters were ligated
to the blunt-ended cDNA fragments followed by XhoI
digestion. The cDNA fragments were directionally cloned
into the EcoRI-XhoI restriction site of the pBluescript
vector. The ligated cDNA fragments were transformed into
DH10B host cells (Gibco BRL). This library was constructed
by M. Bhattacharyya in the laboratory of Dr. Randy
Shoemaker at Iowa State University."
```

```

RESULT 12
AF029423      AF029423      925 bp      DNA      linear      GSS 29-AUG-2000
LOCUS         AF029423      Salmonella typhimurium LT2, Lambda DASH II salmonella
DEFINITION    cyphimurium genomic clone 1244-T3, genomic survey sequence.
ACCESSION     AF029423
VERSION       AF029423.1      GI:2570953
KEYWORDS      GSS.
SOURCE        Salmonella typhimurium
ORGANISM      Salmonella typhimurium
Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
Enterobacteriaceae; Salmonella.
REFERENCE     1 (bases 1 to 925)
AUTHORS       Wong,R.M.-Y., Wong,K.K., Benson,N.R. and McClelland,M.
TITLE         Sample sequencing of a Salmonella typhimurium LT2 lambda library;
              comparison to the Escherichia coli K12 genome
JOURNAL       FEMS Microbiol. Lett. 173 (2), 411-423 (1999)
MEDLINE       99243757
PUBMED        10227170
COMMENT       Contact: McClelland M
              Molecular Biology
              Sidney Kimmel Cancer Center
              3099 Science Park Road, San Diego, CA 92121, USA
              Email: mclelland@lifsc1.sdsu.edu
              Class: shotgun.
FEATURES             Location/Qualifiers
     source          1..925
                     /organism="Salmonella typhimurium"
                     /mol_type="genomic DNA"
                     /strain="LT2"
                     /db_xref="taxon:602"
                     /clone="1244-T3"
                     /clone_lib="Salmonella typhimurium LT2, Lambda DASH II"
                     /note="Vector: Lambda DASH II; sequenced using Li-Cor
                     sequencer."
ORIGIN
Query Match      6.3%; Score 78.2; DB 28; Length 925;
Best Local Similarity 48.4%; Pred. No. 1.4e-05;
Matches 281; Conservative 0; Mismatches 293; Indels 7; Gaps 2;

QY 177 ACTCGGAACATCATCGGCCGACGTGAAGGCACTGATCGGAGCTTCCTGCGATCGCGT 236
Db 50 AGCTCGAAATTTATATGCCGCCCTTTGGCGCACAAATTTCCGGAGCAGGTGATTTTAAG 109
QY 237 CGGTTTCACACTTCGATTCGCGAAACGGCGGGATGTTTGATGCACTGCAGCGGTGT 296
Db 110 CGGTTTCGATATATGATACGCTCGTCAACGGCGGCAATCTGGACGGCAATTCGGCGCGCT 169
QY 297 GTGCGCCCTTTGAGGCTGCGCGGCTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATT 356
Db 170 TGCGCGCTGCTGGCGCTCGACTGGCTAAAGCGACCTACGGCGCGCGCTCSACC-GT 228
QY 357 TGAGTTTCATCGGATCGTGGAGAGAGAGGGCGCGCTTCACAGTGGCATTTGGCGGG 416
Db 229 CGAGGTAGTGTCAATAGGCTTGAGAGAGAGGCGCGCTTTCCCTTACGTTTTTTGGGGCAG 288
QY 417 CGGGGCCATTGACGGGTTGGTCGCCGACAGGAACTGGACTCTTTGGTTGATGAGATCG 476
Db 289 CAAGAACATTTTGGCTGGCGGAATCCCTGAAGAGGTACGCCATATCCAGGAGCCCAAGG 348
QY 477 AGTGTCCGTTAGCGAGCGGCTACTGCTCTTCGGCTTGAAGCGCGGCAACTCAGGCTGC 536
Db 349 CACAGGCTTTGTCGAGCTATCAGCGCGTGCAGATTACCTTCCTCCGCGCGCGCTCGC 408
QY 537 AGCCCGCTCCGGCGGCGGACCTCGGTGCTTTTATCGAACTACACATTGAACAGGACCGAT 596
Db 409 TCGCGGAACC-----GATATCCGGCGGCTTTGTTGAATTCGATATTTGACAGGCTGCGT 462
QY 597 CCTCAGGAGGAGCAATAGATCGAGTTTACCTCCATCGTTGGGTTTCGCGCATT 656

```

```

Db 463 TCTGAGAGCAACGAGCAGTCGATTTGGCGTCGTCAACGCCATCGTCGGGCAACGTCGCTA 522
QY 657 GCGGTTTCCGCTCAAGGAGCAGAGCGACCAACCGCGGCAACACCCCATCATCGTCGCGCA 716
Db 523 CACGCTGACGCTCAACGGCGAGTCCAAACCATGACAGGCAACGCGCAATGGGCTATCGTCG 582
QY 717 GGATCGCGTGTATACCGCGCGCTCTCATGTTGTTGAGGAGGTCA 757
Db 583 CGATACGGTGACCGCCTTCAGTCGCTATTTTTCAGTCAGTCAGTCCA 623

RESULT 13
BI498910      LOCUS
DEFINITION    sai24c10.y1 Gm-cl053 Glycine max cDNA clone GENOME SYSTEMS CLONE
              ID: Gm-cl053-3980 5' similar to TR:085664 085664 N-CARBAMYL-L-AMINO
              ACID AMIDOHYDROLASE. ;, mRNA sequence.
ACCESSION     BI498910
VERSION       BI498910.1      GI:15338254
KEYWORDS      EST.
SOURCE        Glycine max (soybean)
ORGANISM      Glycine max
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae;
Glycine.
REFERENCE     1 (bases 1 to 514)
AUTHORS       Shoemaker,R., Keim,P., Vodkin,L., Erpelting,J., Coryell,V.,
              Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J.,
              Beck,C., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M.,
              Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N.,
              Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M.,
              McCann,R., Waterston,R. and Wilson,R.
              Public Soybean EST Project
              Unpublished (1999)
              Contact: Shoemaker R/Public Soybean EST Project
              Public Soybean EST Project
              Washington University School of Medicine
              4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
              Tel: 314 286 1800
              Fax: 314 286 1810
              Email: est@watson.wustl.edu
              This clone is available through: ResGen, Invitrogen Corp. 2130
              South Memorial Parkway Huntsville, AL 35801 For further information
              call: (800)-533-4363 or contact via email: ccu@resgen.com
              High quality sequence stop: 419.
FEATURES             Location/Qualifiers
     source          1..514
                     /organism="Glycine max"
                     /mol_type="mRNA"
                     /db_xref="taxon:3847"
                     /clone="GENOME SYSTEMS CLONE ID: Gm-cl053-3980"
                     /tissue_type="Whole seedling, 3 week old, greenhouse
                     grown"
                     /lab_host="DH10B"
                     /clone_lib="Gm-cl053"
                     /note="Vector: pBluescript II SK+; Site_1: EcoRI; Site_2:
                     XhoI; The Harosoy NIL was constructed and seed was
                     provided by Dr. J. Specht, University of Nebraska
                     (Shoemaker and Specht, 1995). The cDNA library was
                     constructed from mRNA isolated from whole seedlings of 3
                     week old greenhouse grown plants. Complementary DNA was
                     synthesized from mRNA using a primer consisting of a
                     poly(dT) sequence with a XhoI restriction site and a 3'
                     anchor. EcoRI adapters were ligated to the blunt-ended
                     cDNA fragments followed by XhoI digestion. The cDNA
                     fragments were directionally cloned into the EcoRI-XhoI
                     restriction site of the pBluescript vector. The ligated
                     cDNA fragments were transformed into DH10B host cells
                     (GibcoBRL). This library was constructed in cooperation
                     with Dr. Paul Keim's laboratory at Northern Arizona
                     University."
ORIGIN

```

```
ACCESSION B2566026
VERSION B2566026.1 GI:27194755
KEYWORDS
SOURCE
ORGANISM Pseudomonas aeruginosa
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE 1 (bases 1 to 885)
AUTHORS Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL J. Bacteriol. (2002) In press
COMMENT Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
FEATURES             Location/Qualifiers
     source            1..885
                     /organism="Pseudomonas aeruginosa"
                     /mol_type="genomic DNA"
                     /strain="2-164"
                     /db_xref="taxon:287"
                     /clone="pacs2-164_587"
                     /clone_lib="pacs2-164"
                     /notes="clinical isolate 2-164 Whole genomic shotgun
                     library."
ORIGIN
Query Match      6.8%; Score 83.8; DB 28; Length 885;
Best Local Similarity 51.4%; Pred. No. 9.6e-07;
Matches 242; Conservative 0; Mismatches 227; Indels 2; Gaps 2;
QY 686  ACGCGGACACACCCCGTACCTGGCCAGATGGGTACCGCGCTCTCATGG 745
Db 520  ACGCGGGCAACCGAATGCACTTCCCAAGGACCCCTGGTGGTGGCCCGCTTTG 461
QY 746  TGAGGAGGTCAACCGGTTCGTCAACAGAGATCCCGATGGCACAGTG-GCTACCGTTGGC 804
Db 460  TCCAAGGTTCAATCGGGAAGCCCTCGGCCAATCAGCCGATGGTTGGCGCAGGTGCGA 401
QY 805  CACCTCACAGTGCCCGCGGTGAGCAACCAAGTCCCGGGGAGGTGCACTTCACTG 864
Db 400  TGCCTGCAGCCCTATCCCGTTCCGCAACGTCATACCCCGCAAGTGAAGATGACCCCTG 341
QY 865  GACCTGGTTCTCCGATGAGGAGTCGTCGCGTCTGATCGACCGCATCTCGGTCATG 924
Db 340  GACTTCGCGCATCTGCAACCGGAGCGCCCTGGACTCGATGATCGCGAGGTCCGCGCAGGTG 281
QY 925  GTCCGGAGGTGCCTCCCGAGCGCGGTGGTGGCTGCCGATGTGGATGAATTTTCAATCTC 984
Db 280  ATGCGCGCTACTGCGAGAGAGATGCTTGCAATACAGCTGGTGGCGACCGCGATTTTC 221
QY 985  AGCCCGGTGAGTGCTCCTACATGTGGAGCGCGTTTCGCGAAGCGCCTCGGCGTTG 1044
Db 220  CGCGCGCTACTTCCAGCAGGAGTCCGTCCGGC-CGGTGGCGAGCGCGCAGGCGGTG 162
QY 1045  CAGTTACACACCGGATATCAGATGGGCGGGCCAGCACTCGATGTTCAATCGCCAG 1104
Db 161  GGCATCCCGAGATGACATCGTCAGCGCGCGCGCCACGACGCGATCTTCTCGCCGAA 102
QY 1105  GTCACGAGCTCGGAATGGTTTCGTTTCCAAGCGGTGCTGGCGGAGCCAC 1155
Db 101  CTCGGTCCGCGGGAGTATCTTCGTGCCCTGGGAAACGGCATCAGCCAC 51
RESULT 11
CD443670
LOCUS
```

```
DEFINITION EL01N0429G06.b Endosperm_4 Zea mays cDNA, mRNA sequence.
ACCESSION CD443670
VERSION CD443670.1 GI:31359313
KEYWORDS EST.
SOURCE Zea mays
ORGANISM Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
REFERENCE 1 (bases 1 to 572)
AUTHORS Lai,J., Dey,N., Kim,C.S., Becraft,P., Larkins,B., Linton,E. and
Messing,J.
TITLE Sequencing of the maize endosperm ESTs
JOURNAL Unpublished (2002)
COMMENT Contact: Lai, Jinsheng
Dr. Joachim Messing's lab
Waksman Institute, Rutgers University
190 Frelinghuysen Rd., Piscataway, NJ 08854, USA
Tel: 732-445-3801
Fax: 732-445-5735
Email: jlai@waksman.rutgers.edu
Seq primer: T3.
FEATURES             Location/Qualifiers
     source            1..572
                     /organism="Zea mays"
                     /mol_type="mRNA"
                     /cultivar="W22"
                     /db_xref="taxon:4577"
                     /tissue_type="Endosperm of 7-23DAP"
                     /clone_lib="Endosperm 4"
                     /note="Vector: pBluescript SK-; Site_1: EcoRI; Site_2:
                     XhoI"
ORIGIN
Query Match      6.4%; Score 79.2; DB 14; Length 572;
Best Local Similarity 49.8%; Pred. No. 7.1e-06;
Matches 260; Conservative 0; Mismatches 253; Indels 9; Gaps 2;
QY 226  GCGATCGCGGTGCGTTCACTTCGATTCGTCCGAAACGGCGGATGTTGTATGGCACT 285
Db 13  GCAGTTGCCACTGGATCTCATGTGTGTCATTCCTCAGGCAAAATTCATGGTGT 72
QY 286  GCAGCGGTGCTGGCGCCTTCAGCTGCCGGTGATCTGGAGAACGCTACGTGGAAT 345
Db 73  GTTGAGTCTCGGTGCTCTTCAGGCAATTAGTTTCTGAAAAGGTCTGCTTCTCTGCCA 132
QY 346  CGGCATCCATTTGAGTTTCATCGCATCGTGGAGGAGAAAGGGCCCGCTTCAGCAGTGGC 405
Db 133  AAAAGATCTTTGAGGTTATTATGTTTACATCAGAGGAGCGCACAGATTGGAAATTAGC 192
QY 406  ATGTTGGCGCGCGGCCCATTCGAGGTTGGTCGCCGACAGGGAA---CTGACTCTTTG 462
Db 193  TCGTTGGGAAGCGCGCTTAATGGCAGGATTGAAGAATCTTGCTCAATCCCTCCGAAAGTA 252
QY 463  GTTGATGAGGATGGAGTGCCTTAGGCAGGCGG-----CTACTGCCCTTCGGCTTGAAG 516
Db 253  GTTGACAATCAGAAATGTGTCATTTTGGATGCTGCAAGAAATCTGCTGGCTAATAGTTGAAT 312
QY 517  CGCGGCGCACTCGAGCTGCACCCGCTCCGCGCGGAGCTCGCTGCTCTTTATCGAACTA 576
Db 313  CTAGAGGACCTCATAGCGTATTTTAAAGACAGACAAATACTCTGCTTTTATAGAACTG 372
QY 577  CACATTGAACAAGGACCGATCTCGAGCAGAGAGCAAAATAGAGATCGAGTTGTAACTCC 636
Db 373  CACATTGAGCAGGTCCTCATCTTAGAAAGAGGAGGTATCCCTATTGGCATTTGTTACTGCA 432
QY 637  ATCGTTGGGTTCCGCGCATTCGGGTTCGCGTCAAGGCAGAGGAGGACACCGCCGAC 696
Db 433  ATTGCTGCTCTGCAAGTCTTAAGGTGGACTTTGAAGGGAATGGCGGTCTATGCTGGAGCA 492
QY 697  ACCCCCATGCACCTCGCGCAGGATCGCTGGTGTACCCGCGCT 738
Db 493  GTGCTGATGCTCGCAAGAAATGATGCTGGACTGGCAGCAGCT 534
```


Best Local Similarity 48.8%; Pred. No. 8.5e-09;
Matches 292; Conservative 0; Mismatches 294; Indels 12; Gaps 1;
QY 146 AGCGGCGGCTTGAGGGTTCTGTAAGACGCACTCGGAACATCATCGGCGGACGTGAAG 205
DB 68 AGCGGCGGCGCTTCAGATGGGCGTCAAGAGATGGGCACGATGTTGCGCACCGCCCG 127
QY 206 GCACCTGATCGGAGCTTCCTGCGCATCGCGTCTGGTTTCACACTTCGATTCTGTCGAAACG 265
DB 128 GCGAGACCGGATGACATGCGCGGTCTATATCGGAGCCATCTCGACCCAGCCACCG 187
QY 266 GCGGGATGTTGATGCACTGCAAGGCGTGTGCGCCCTTGAGGCTGCGCGGATGATGC 325
DB 188 GCGGCAAGTTCGATGCGGCTTCGCGGTGCTGGCTGCACTGGAAGTCGTGCGAAGCCTGA 247
QY 326 TGGAGAACGGCTACGTGAATCGGCATCCATTGAGTTTCATCGGCATCGTGGAGGAGGAAG 385
DB 248 ACGATCTCAACATAAAGACGAAACATTTTTTGTGCTGTTACCAACTGGTCCAACTAGGAAG 307
QY 386 GGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGGCCCATTTGCAAGGTTGGTTCGCGGACA 445
DB 308 GCGCTCGCTTTGCGCGGCGCATGCTGCTTCGCGGTCTTCGCGGCAATTCAGATCTCG 367
QY 446 GGGAACTGCACTCTTTGTTGATGAGGATGGAGTGTCCGTTAGGCGAGCGGCTACTGCGCT 505
DB 368 ATTACGCTATAGCCCGCATACCGGACGAGCAACCTATGCGGACGAACCTGAAGCGCA 427
QY 506 TCGGCTTGAAGCGCGGCGCACTCGAGCTGCGACCCGCTCCGCGCGGACCTCGCTGCTT 565
DB 428 TCGGCTTGGCTGGTGAAGAGAGTGGCGCGCCG-----AAGATGCAAGCCT 475
QY 566 TTATCAACTACATGTAACAGGACCGATCTCGAGCAGGAGCAATAGATAGATCGGAG 625
DB 476 ATTTCGAATATCATATCGAACAGGACCGCATCTCGAGGCGGAAGCAAGCATCGCGG 535
QY 626 TTGTAACTCCATCGTTGCGCGCATTTGCGGCTTGCCTGTCGCAAGGCGAAGCGAC 685
DB 536 TCGTTACCATGTCAGGCGCTGTGTGCTGGAAGTGACCTGACGGGCAAGGAGCGC 595
QY 686 AGCGCGGCAACCCCGCATCGACCTGCGGCGAGATCGGCTGTAACCGCGCTCTCAT 743
DB 596 ATACCGGCTCGAGCGCGATGGCCATCGCGTCAATGCGCGCTTCGCGCGCGCGCGCAT 653

RESULT 7
BZ576481/c
LOCUS
DEFINITION
msh2_4953.x1 msh Pseudomonas aeruginosa genomic clone msh2_4953,
genomic survey sequence.
ACCESSION
BZ576481
VERSION
BZ576481.1 GI:27211542
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1522)
AUTHORS
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
LOCATION/Qualifiers
1. .1522
/organism="Pseudomonas aeruginosa"

/mol_type="genomic DNA"
/strain="W5H"
/db_xref="taxon:287"
/clone="msh2_4953"
/notes="Environmental isolate. Whole genomic shotgun
library."

ORIGIN

Query Match 7.5%; Score 93.4; DB 28; Length 1522;
Best Local Similarity 52.4%; Pred. No. 1.3e-08;
Matches 205; Conservative 0; Mismatches 186; Indels 0; Gaps 0;
QY 577 CACATTGAACAAGGACCGATCTCGAGCAGGAGCAATAGAGATCGGAGTTGTAACTCC 636
DB 511 CACATCAACAGGGCGCATCTCGAGGACGAGGAGACCATAGCGTGGTGTGCGG 452
QY 637 ATCTGTTGGGTTTCGCGCATTTGCGGTTGCGTCAAAAGGAGAGACGACACCGCGGACA 696
DB 451 GCGCTCGGCGCAGAAATGGTTTCGACCTGTCCTTGCACGCGCTCGAGGCACACGCGGCGCA 392
QY 697 ACCCCATGCACTGCGGCGAGATGCGTGGTACCGCGCGCTCTCATGTTGAGGAGGTC 756
DB 391 ACGCCGATGACCTTGCCTAAGGACGCGCTGCGTGGTCCGCGCGGTCGAGGCGGTC 332
QY 757 AACCGTTTCGTCACAGAGATGCGCGATGCGCATGCGTACCGTTGCGCCACCTCACAGTG 816
DB 331 AATCGCGCGGCGCTTCGCGCATCAGCGCGATGCTTTCGCGCACGCTCGCTGCGCGCC 272
QY 817 GCCCGCGTGGAGGCAACACAGGTCCCGGGGAGGTGGACTTCACACTGGACCTGCGTTCT 876
DB 271 TATCCGCTTTCGCGCAACGTCGATACCGCGGAGTGAAGTACACCTGGACTTCCGCCAT 212
QY 877 CCGCATGAGAGTCGCTCCGCTGCTGATCGACCGCATCTCGGTCATGTTGCGGAGGTC 936
DB 211 CTGCAACCGGAGCGCTGCGACTCGATGATCGCGAAGTCCGCCAGGTGTCGCGCTTACC 152
QY 937 GCTTCCCGAGGCGGTGCTGCTGCGCGATGCG 967
DB 151 TCGGAGAACATGGCTTGCAATACGAGCTGG 121

RESULT 8

BZ549118/c
LOCUS
DEFINITION
paci-60_1678.s1 paci-60 Pseudomonas aeruginosa genomic clone
paci-60_1678, genomic survey sequence.
ACCESSION
BZ549118
VERSION
BZ549118.1 GI:27152699
KEYWORDS
GSS.
SOURCE
Pseudomonas aeruginosa
ORGANISM
Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
REFERENCE
1 (bases 1 to 1250)
AUTHORS
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
TITLE
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
JOURNAL
J. Bacteriol. (2002) In press
COMMENT
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
LOCATION/Qualifiers
1. .1250
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="1-60"

FEATURES
source

```

Db      389  CCAGGATCGTCGGCGCGTGCAGCGCGCGAGCGCGTGGCGATCCGACAGATGA 448
Qy      1062 TATCAGCACTGGGCGCGCCAGCACTCGATGTTTCAGCGCCAGGTACAGGCG-TCGGAA 1120
Db      449  CATCGTCAGCGCGCGCGCGCAGCAGCGATCTTCCTCGCGCAACTCGGTCCGCGGGGA 508
Qy      1121 TGGTTTTCGTTTCAAGCCGTGTGCGCGGAGCAGCATGTTCCCGAAGAAATGGACCGATTTTCG 1180
Db      509  TGATCTTCGTGCTCGGAGACGGCATCAGCCACAGAGATCGAAGACCGCGCCCG 568
Qy      1181 ATGACCT 1187
Db      569  AGGACCT 575

RESULT 5
BZ570771/c
LOCUS      1343 bp      DNA      linear      GSS 17-DEC-2002
DEFINITION msh2_1528.y2 msh Pseudomonas aeruginosa genomic clone msh2_1528,
            genomic survey sequence.
ACCESSION  BZ570771
VERSION    BZ570771
KEYWORDS  BZ570771.1 GI:27205832
SOURCE    GSS.
ORGANISM  Pseudomonas aeruginosa
            Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
            Pseudomonadaceae; Pseudomonas.
REFERENCE  1 (bases 1 to 1343)
            Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
            Burns,J.L., Kaul,R. and Olsen,M.V.
            Whole-Genome-Sequence variation among multiple isolates of
            Pseudomonas aeruginosa library
            J. Bacteriol. (2002) In press
            Contact: Chris K. Raymond
            Genome Center
            University of Washington
            Box 352145, Seattle, WA 98105-2145, USA
            Tel: 2062216954
            Fax: 2066857244
            Email: craymond@u.washington.edu
            Class: shotgun.
FEATURES   Location/Qualifiers
            source
            1..1343
                /organism="Pseudomonas aeruginosa"
                /mol_type="genomic DNA"
                /strain="MSh"
                /db_xref="taxon:287"
                /clone="msh2_1528"
                /clone_lib="msh"
                /note="Environmental isolate. Whole genomic shotgun
                library."
ORIGIN
Query Match      7.7%; Score 95.4; DB 28; Length 1343;
Best Local Similarity 49.4%; Pred. No. 4.8e-09;
Matches 312; Conservative 0; Mismatches 306; Indels 13; Gaps 2;

Qy      263  ACGCGGGATGTTTGTATGCACTGCAGGCGTGTGCGCCCTTGAGGTCGCGCGTGA 322
Db      699  ACCGTTGCAAGTTCGACGCTGTTTCGGGGTGATGCCCGCTCGAGGTGATCCGSCACC 640
Qy      323  TGTGTGAGAACGCTACGTGAATCGGCATCTTCAGTTTCATCGCGATCGTGGAGAGG 382
Db      639  CTCACGAACTCGGGTGGAAACCCAGCGCGCTGAAGTGTGTGTGACACCGAG 580
Qy      383  AAGGGCCCGCTTACAGCACTGGCATGTTGGCGCGCGGCCCAATTCAGGAGTTGGTGGCG 442
Db      579  AAGGCTCGCGCTTCGCGCCCTGCATGATGGGCTCGGCGTATTTCGCGGGAAGTTCACCC 520
Qy      443  ACAGGAACTGGAAGTCTTTGGTTGATGAGGATGGAGTCCGTTAGGACGGCGGCTACTG 502
Db      519  TGAGGAGAC-CCTGGCCAAAGCGCGATGCCGACGGTGTACGCGTAGGCGAGGCGGTGAGC 461

```

```

Qy      503  CTTTCGGCTTGAAGCGCGCGAACTGCAGGCTGCAGCCGCTCCCGCGCGGACCTCGCTG 562
Db      460  CCATCGGT-----ACGCGGAGCGCGGATGTTCTCGACATCCGGTGGCG 413
Qy      563  CTTTATCGAACTACACATTGAACAGGACCGATCTCGAGCAGGAGCAAAATAGAGATCG 622
Db      412  CTAATTCGAGCGCACATCGACAGGCGCGATCCTCGAGGACGAGGAGAACCATCG 353
Qy      623  GAGTTGTAACCTCCATCGTTGGCGTTCCGCATTCGGGTTGCCGTCAAAGCAGAGCG 682
Db      352  GCGTGGTGTCTCGGCGCGCTCGGGCGAGAAATGGTTGCACTGTCTCGCGCGCTCGAGG 293
Qy      683  ACCACGCGCGCACAAACCCCATGCACTGCGCGAGGATCGCTGTATACCCCGCGCTCTCA 742
Db      292  CACAGCGCGCGCCAAAGCGCGATGCACTGCGCAGGACGCCCTGGTCGGTCCCGCGCG 233
Qy      743  TGGTGAAGGAGGTCAACCGTTTCGTCACAGAGATCGCGATGGCAACAGTGGTACCGTTG 802
Db      232  TGGTCGAGCGGTCAATCGCGCGCTTCGCGCATCAGCCGCGATGCTTCGCGCACGGTCG 173
Qy      803  GCCACCTCACAGTGGCCCCCGGTGGAGCAACCAAGTCCCGGGGAGGTGAGCTTCACAC 862
Db      172  GCTGCTGCACGCTATCCCGGTTTCGCGCAACGTATACCCCGCGAAGTGAAGATGACCC 113
Qy      863  TGGACTCGCTTCTCCGCATGAGGAGTCGCT 893
Db      112  TGGACTTCGCCCATCTGCAGGGGATCCACT 82

CC130324      694 bp      DNA      linear      GSS 16-APR-2003
NDL.58C11.SP6 Notre Dame Liverpool Aedes aegypti genomic clone
CC130324      ND.L.58C11, genomic survey sequence.
GSS.          CC130324.1 GI:29999379
SOURCE        Aedes aegypti (yellow fever mosquito)
ORGANISM      Aedes aegypti
            Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
            Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Aedes;
            Stegomyia.
REFERENCE  1 (bases 1 to 694)
            Loftus,B., Shetty,J., Knudson,D. and Severson,D.
            BAC end sequencing of Aedes aegypti
            Unpublished (2003)
            Other_GSSs: NDL.58C11.77
            Contact: Brendan Loftus
            Department of Eukaryotic Genomics
            TIGR
            9712 Medical Center Drive, Rockville, MD 20850, USA
            Tel: 301-838-3543
            Fax: 301-838-0208
            Email: enta@tigr.org
            Library was provided by David Severson
            Seq primer: SP6
            Class: BAC ends.
FEATURES   Location/Qualifiers
            source
            1..694
                /organism="Aedes aegypti"
                /mol_type="genomic DNA"
                /strain="liverpool"
                /db_xref="taxon:7159"
                /clone="NDL.58C11"
                /clone_lib="Notre Dame Liverpool"
                /note="Vector: pECBAC1; Site 1: Hind III; The library was
                prepared from whole body tissue of newly hatched L1 larvae
                by David Severson at the University of Notre Dame and
                Hongbin Zhang"
ORIGIN
Query Match      7.6%; Score 93.6; DB 28; Length 694;

```


GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model
Run on: May 4, 2004, 13:15:23 ; Search time 1685 Seconds
(without alignments)
10040.500 Million cell updates/sec

Title: US-10-045-063-1
Perfect score: 1239
Sequence: 1 gtgacctgcagaagcgca.....tgaaggcacttgaccggtaa 1239

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 27513289 seqs, 14931090276 residues
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_htc:*
9: gb_esti:*
10: gb_est2:*
11: gb_htc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vri:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	125.8	10.2	935	28	BZ549119 pacel1-60_
c 2	96.6	7.8	614	28	BZ349536 hr41h12.g
c 3	95.8	7.7	667	14	CA757386 BR0500120
4	95.8	7.7	1226	28	BZ567409 pacel2-164

c 5	95.4	7.7	1343	28	BZ570771	BZ570771 msh2 1528
6	93.6	7.6	694	28	CC130324	CC130324 NDL_58C11
c 7	93.4	7.5	1522	28	BZ576481	BZ576481 mah2_4953
c 8	92.8	7.5	1250	28	BZ549118	BZ549118 pacel1-60
9	90.6	7.3	642	28	BZ242460	BZ242460 id54a10.15
c 10	83.8	6.8	885	28	BZ566026	BZ566026 pacel2-164
11	79.2	6.4	572	14	CD443670	CD443670 EL01N0429
12	78.2	6.3	925	28	AF029423	AF029423 AF029423
13	76.6	6.2	514	12	BI498810	BI498810 sai24c10.
14	76.6	6.2	565	12	BI973078	BI973078 sag86f04.
15	76	6.1	669	13	BX255493	BX255493 BX255493
c 16	75.4	6.1	1285	28	BZ556662	BZ556662 pacel1-60
17	75	6.1	543	14	CA202312	CA202312 SCRUF101
18	74.6	6.0	542	14	CF603479	CF603479 BACCA01.0
19	74.6	6.0	551	12	BZ200157	BZ200157 BZ200157
20	74.2	6.0	495	10	BE602857	BE602857 HVSMEN010
21	74.2	6.0	547	12	BJ478213	BJ478213 BJ478213
22	74.2	6.0	752	14	CB670856	CB670856 OSJNRe04C
23	73.4	5.9	542	12	BI419934	BI419934 LJNEST42a
24	73.4	5.9	822	14	CD445974	CD445974 EL01T0206
25	73	5.9	656	14	CB005475	CB005475 VVC022C10
26	72.6	5.9	474	10	BE360524	BE360524 DGI_64.G0
27	72.6	5.9	624	12	BF184695	BF184695 BF184695
28	71.2	5.7	718	13	BQ996682	BQ996682 QGG13103.
29	70.8	5.7	862	28	BZ556335	BZ556335 pacel1-60
c 30	70.4	5.7	1085	28	BZ557114	BZ557114 pacel1-60
c 31	69.8	5.6	781	12	BZ582339	BZ582339 BZ582339
32	69.6	5.6	658	14	CF451325	CF451325 EST687670
33	69.4	5.6	870	10	BF626992	BF626992 HVSMEN000
34	68.6	5.5	596	13	BU014375	BU014375 QGJ7C21.Y
35	68.4	5.5	634	10	AW573986	AW573986 EST316577
36	68.4	5.5	685	12	BI308146	BI308146 EST529556
37	67.8	5.5	443	14	CF324522	CF324522 HDN--06.M
38	67.4	5.4	644	13	BQ148538	BQ148538 NF069D02F
39	67	5.4	799	13	BU001543	BU001543 QGG28C05.
40	66.8	5.4	770	14	CB672554	CB672554 OSJNRe06K
41	66.6	5.4	771	9	AU088788	AU088788 AU088788
42	65.4	5.3	613	13	BQ591711	BQ591711 E012617-0
43	65.4	5.3	846	14	CF652670	CF652670 69-L02057
44	65.2	5.3	630	10	AW585686	AW585686 EST317309
45	65	5.2	816	14	CB644200	CB644200 OSJNRe05E

ALIGNMENTS

RESULT 1
BZ549119/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BZ549119 935 bp DNA linear GSS 17-DEC-2002
pacel1-60_1678.s2 pacel1-60 Pseudomonas aeruginosa genomic clone
pacel1-60_1678, genomic survey sequence.

BZ549119 GI:27152700

Pseudomonas aeruginosa

Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

Pseudomonadaceae; Pseudomonas.

1 (bases 1 to 935)

Spencer, D.H., Raymond, C.K., Smith, E.E., Sims, E.E., Hastings, M.,

Burns, J.L., Kaul, R. and Olsen, M.V.

Whole-Genome-Sequence variation among multiple isolates of

Pseudomonas aeruginosa library

J. Bacteriol. (2002) In press

Contact: Chris K. Raymond

Genome Center

University of Washington

Box 352145, Seattle, WA 98105-2145, USA

Tel: 20622216934

Fax: 2066857244

Email: craymond@u.washington.edu

Class: shotgun.

Location/Qualifiers

FEATURES

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 12:20:58 ; Search time 5119 Seconds
(without alignments)
10490.725 Million cell updates/sec

Title: US-10-045-063-1
Perfect score: 1239
Sequence: 1 gtgacccctgcgaagcgca.....tgaaggcacttgaccggttaa 1239

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3470272 seqs, 21671516995 residues
Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sta.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_mu.*
- 20: em_om.*
- 21: em_or.*
- 22: em_ov.*
- 23: em_pat.*
- 24: em_ph.*
- 25: em_pl.*
- 26: em_ro.*
- 27: em_sta.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htg_hum.*
- 31: em_htg_inv.*
- 32: em_htg_other.*
- 33: em_htg_mus.*
- 34: em_htg_pln.*
- 35: em_htg_rod.*
- 36: em_htg_mam.*
- 37: em_htg_vrt.*
- 38: em_ey.*
- 39: em_htgo_hum.*
- 40: em_htgo_mus.*
- 41: em_htgo_other.*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1239	100.0	1239	6	AR198375	Sequence
2	1239	100.0	1239	6	AX015596	Sequence
3	1239	100.0	1239	6	BD194451	Arthrobac
4	1238	99.9	1239	6	AX103678	Sequence
5	1238	99.9	1239	6	AX370705	Sequence
6	1237.4	99.9	8715	1	AF146701	Arthrobac
7	1192.4	96.2	1263	6	AX370708	Sequence
8	739.8	59.7	1239	6	AX397881	Sequence
9	739.8	59.7	1239	6	BD176614	DNA encod
10	739.8	59.7	1239	6	BD180889	DNA encod
11	739.8	59.7	1239	6	BD181025	5-Substit
12	739.8	59.7	3343	6	AX397883	Sequence
13	739.8	59.7	3343	6	BD176615	DNA encod
14	739.8	59.7	3343	6	BD180890	DNA encod
15	739.8	59.7	3343	6	BD181026	5-Substit
16	235.8	19.0	280050	1	AL591975	Listeria
17	235.8	19.0	349980	6	AX641665	Sequence
18	225.2	18.2	299910	1	AP005957	Bradyrhiz
19	222.6	18.0	11322	1	AE012124	xanthomon
20	220.6	17.8	249050	1	AL596165	Listeria
21	220.6	17.8	349980	6	AX417038	Sequence
22	220.6	17.8	349980	6	AX417041	Sequence
23	214.8	17.3	11276	1	AE004481	Pseudomon
24	210.8	17.0	1380	6	AR385713	Sequence
25	207.6	16.8	13275	1	AE011655	xanthomon
26	205.6	16.6	348997	1	BX640427	Bordetell
27	205.6	16.6	349008	1	BX640444	Bordetell
28	194.4	15.7	10813	1	AE005928	Caulobact
29	193	15.6	304517	1	AE016789	Pseudomon
30	191.4	15.4	346547	1	AP003012	Mesorhizo
31	190.8	15.4	348706	1	BX640445	Bordetell
32	190.8	15.4	349354	1	BX640416	Bordetell
33	186.8	15.1	299991	1	AE016776	Pseudomon
34	186.4	15.0	300750	1	AP001509	Bacillus
35	183.8	14.8	1582	1	AF425838	Geobacill
36	182.2	14.7	1860	6	E03775	DNA sequenc
37	182.2	14.7	2731	1	S67784	N-carbamyl-
38	181.8	14.7	329709	1	AP002997	Mesorhizo
39	181.8	14.7	346362	1	BX640439	Bordetell
40	181.8	14.7	348257	1	BX640425	Bordetell
41	179	14.4	1577	1	BSAMAB	B.stearothe
42	179	14.4	4839	1	BSORF4	B.stearothe
43	171.8	13.9	299350	1	AP004604	Oceanobac
44	171.2	13.8	348934	1	BX640417	Bordetell
45	170.8	13.8	344615	1	BX569695	Synechoco

ALIGNMENTS

RESULT 1	AR198375	Sequence 1	1239 bp	DNA	linear	PAT 20-APR-2002
LOCUS	AR198375	Sequence 1	from patent US 6352848.			
DEFINITION	AR198375					
ACCESSION	AR198375.1					
VERSION	AR198375.1					
KEYWORDS						
SOURCE	Unknown.					
ORGANISM	Unknown.					

REFERENCE 1 (bases 1 to 1239)
AUTHORS Altenbuchner,J., Mattes,R., Pietzsch,M., Syldack,C., Wiese,A. and Wilms,B.
TITLE Recombinant L-N-carbamoylase from Arthrobacter aureusens and method of producing L-amino acids therewith

JOURNAL Patent: US 6352848-A 1 05-MAR-2002;
FEATURES Location/Qualifiers
source 1..1239
/organism="unknown"
/mol_type="unassigned DNA"

ORIGIN

Query Match 100.0%; Score 1239; DB 6; Length 1239;
Best Local Similarity 100.0%; Pred. No. 8.9e-208;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAGAAAGAGATCCGGAGCTCTCCCGG 60
Db 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAGAAAGAGATCCGGAGCTCTCCCGG 60

Qy 61 TTCTCGGAGAAAGCGCGCGGTGTTACCGGCTGACCTACATCCAGAGCATGCCCGCG 120
Db 61 TTCTCGGAGAAAGCGCGCGGTGTTACCGGCTGACCTACATCCAGAGCATGCCCGCG 120

Qy 121 CGGGAAGCGCTATTGCGGCTATGAAAGCGCGCTTGAGCGTTGCTGAAAGCAGCTC 180
Db 121 CGGGAAGCGCTATTGCGGCTATGAAAGCGCGCTTGAGCGTTGCTGAAAGCAGCTC 180

Qy 181 GGAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTCGATCGCGTCTG 240
Db 181 GGAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTCGATCGCGTCTG 240

Qy 241 TCACACTTCGATTCTGTCGAAACGCGGGATGTTTGATGACCTGCGAGCGTGTGTGC 300
Db 241 TCACACTTCGATTCTGTCGAAACGCGGGATGTTTGATGACCTGCGAGCGTGTGTGC 300

Qy 301 GCGCTTAGGCTGCGCGGTGATGCTGAGAAAGCGCTACGTAATCGGATCCATTTGAG 360
Db 301 GCGCTTAGGCTGCGCGGTGATGCTGAGAAAGCGCTACGTAATCGGATCCATTTGAG 360

Qy 361 TTCTCGGATCGTGAGGAGAAAGCGCGCTTTCAGAGTGGCATGTTGGCGCGCGG 420
Db 361 TTCTCGGATCGTGAGGAGAAAGCGCGCTTTCAGAGTGGCATGTTGGCGCGCGG 420

Qy 421 GCATTGAGGCTTGGTCCGACAGGGAAGTGGACTCTTTGGTTGATGAGGATGAGTG 480
Db 421 GCATTGAGGCTTGGTCCGACAGGGAAGTGGACTCTTTGGTTGATGAGGATGAGTG 480

Qy 481 TCGTTAGGAGCGCGCTACTGCTTTCGCTTGAAGCGCGGGAATCGAGGCTGCAGCC 540
Db 481 TCGTTAGGAGCGCGCTACTGCTTTCGCTTGAAGCGCGGGAATCGAGGCTGCAGCC 540

Qy 541 CGCTCGGCGGAGCTCGTGCTTTTATCGAACTACACATTGAACAAGGACCGATCTC 600
Db 541 CGCTCGGCGGAGCTCGTGCTTTTATCGAACTACACATTGAACAAGGACCGATCTC 600

Qy 601 GAGCAGGAGCAATAGAGATCGGAGTGTAACTCCATCGTTGGCGTTTCGGCATTCGG 660
Db 601 GAGCAGGAGCAATAGAGATCGGAGTGTAACTCCATCGTTGGCGTTTCGGCATTCGG 660

Qy 661 GTTCCGCTCAAGGAGAGGAGCAAGCAGCGCGCAAAACCCCATGCTGCGGAGAT 720
Db 661 GTTCCGCTCAAGGAGAGGAGCAAGCAGCGCGCAAAACCCCATGCTGCGGAGAT 720

Qy 721 GCGTGTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCG 780
Db 721 GCGTGTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCG 780

Qy 781 GATGCAAGTGGCTACCGTTGGCCACTTCAAGTGGCGCGCGGTGGAGGCAACAGATC 840
Db 781 GATGCAAGTGGCTACCGTTGGCCACTTCAAGTGGCGCGCGGTGGAGGCAACAGATC 840

Qy 841 CCGGGGAGGTGGAATTACACTGACCTGCTGTTCTCGCATGAGGATCGCTCCGCTG 900
Db 841 CCGGGGAGGTGGAATTACACTGACCTGCTGTTCTCGCATGAGGATCGCTCCGCTG 900

Qy 901 CTGATCGACCGCATCTCGGTCATGCTGCGGAGGTGCGCTCCAGCGCGGTGTGCTGCC 960
Db 901 CTGATCGACCGCATCTCGGTCATGCTGCGGAGGTGCGCTCCAGCGCGGTGTGCTGCC 960

Db 901 CTGATCGACCGCATCTCGGTCATGCTCGGCGAGGTGCGCTCCAGCGCGGTGTGCTGCC 960

Qy 961 GATGTGATGAATTTTCAATCTCAGCCCGTGCAGCTGGCTCTCAATCATGTTGACGCC 1020
Db 961 GATGTGATGAATTTTCAATCTCAGCCCGTGCAGCTGGCTCTCAATCATGTTGACGCC 1020

Qy 1021 GTTCGGAAGCGGCTCGGCTTCAGTTTACACACCGGGATATCAGCAGTGGGCGCGG 1080
Db 1021 GTTCGGAAGCGGCTCGGCTTCAGTTTACACACCGGGATATCAGCAGTGGGCGCGG 1080

Qy 1081 CACGACTCGATGTTTCATCGCCCGAGGTACGAGCGTCCGAAATGTTTTCGTTCCAGCCGT 1140
Db 1081 CACGACTCGATGTTTCATCGCCCGAGGTACGAGCGTCCGAAATGTTTTCGTTCCAGCCGT 1140

Qy 1141 GCTGCGCGAGCCAGTTCGCGAAGATGGAACCATTCGATGACCTTCGCAAGGAACT 1200
Db 1141 GCTGCGCGAGCCAGTTCGCGAAGATGGAACCATTCGATGACCTTCGCAAGGAACT 1200

Qy 1201 GAGGTGTGCTCCGGGTAATGAAGGCATTCGACCGGTAA 1239
Db 1201 GAGGTGTGCTCCGGGTAATGAAGGCATTCGACCGGTAA 1239

RESULT 2
AX015596
LOCUS AX015596 1239 bp DNA linear PAT 07-SEP-2000
DEFINITION Sequence 4 from Patent WO9951722.
ACCESSION AX015596
VERSION AX015596.1 GI:10041431
KEYWORDS
SOURCE Arthrobacter aureusens
ORGANISM Arthrobacter aureusens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococineae; Micrococcaceae; Arthrobacter.

REFERENCE 1
AUTHORS Sylidat,C., Mattes,R., Wilms,B., Pietzsch,M., Wiese,A. and Alrenbuchner, J.
TITLE Recombinant l-n-carbamoylase derived from arthrobacter aureusens, and a method for producing l-amino acids by using the same
JOURNAL Patent: WO 9951722-A 4 14-OCT-1999;
DEGUSA (DE); UNIV STUTTGART (DE); ROCHE DIAGNOSTICS GMBH (DE)

FEATURES
source 1..1239
/organism="Arthrobacter aureusens"
/mol_type="unassigned DNA"
/db_xref="taxon:43663"

CDS
1..1239
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC07595.1"
/db_xref="GI:10041432"
/translation="MTLQKAQAARIKEIRIELRSFSAEGPGVTRLTYPTEHAARETL
IAAMAAALSVREDALGNITGRREGTDPPELPAIAVGHFSVRNGMDFGAVGVVCA
EAAVMLENGVNRPHPEFIAIVEEGARFSSGMLGGRAIAGLVADRLDSLVDGCV
SVROAATAFGLKPELOAARSADLRAFIELHIEQGLEIEIGVTSIVGVRA
LRVAVGRSDHAGTTPMHLKQDALVPAALMVRVFNVEIADGTATVATVHLSIVPGLA
GNQVPGVDFTLRLRSFESRLVIDRIISVMGEVASQAGAAADVFDFNLSVPQLA
PTWVDAVREAAALQFTHRDITSSGAGHDSMFIAQVTDVGMVFVPSRAGRSHVPEWTD
FDDLKRGTEVVLVRVVKALDR"

ORIGIN

Query Match 100.0%; Score 1239; DB 6; Length 1239;
Best Local Similarity 100.0%; Pred. No. 8.9e-208;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GTGACCTTCAGAAAGCGCGCGCATTTGAGAAAGAGATCCGGAGCTCTCCCGG 60
Db 1 GTGACCTTCAGAAAGCGCGCGCATTTGAGAAAGAGATCCGGAGCTCTCCCGG 60

Qy 61 TTCTCGGAGAAAGCGCGCGGTGTTACCGGCTGACCTACATCCAGAGCATGCCCGCG 120
Db 61 TTCTCGGAGAAAGCGCGCGGTGTTACCGGCTGACCTACATCCAGAGCATGCCCGCG 120

Qy 121 CGGGAAGCGCTATTGCGGCTATGAAAGCGCGCTTGAGCGTTGCTGAAAGCAGCTC 180
Db 121 CGGGAAGCGCTATTGCGGCTATGAAAGCGCGCTTGAGCGTTGCTGAAAGCAGCTC 180

Qy 181 GGAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTCGATCGCGTCTG 240
Db 181 GGAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTCGATCGCGTCTG 240

Qy 241 TCACACTTCGATTCTGTCGAAACGCGGGATGTTTGATGACCTGCGAGCGTGTGTGC 300
Db 241 TCACACTTCGATTCTGTCGAAACGCGGGATGTTTGATGACCTGCGAGCGTGTGTGC 300

Qy 301 GCGCTTAGGCTGCGCGGTGATGCTGAGAAAGCGCTACGTAATCGGATCCATTTGAG 360
Db 301 GCGCTTAGGCTGCGCGGTGATGCTGAGAAAGCGCTACGTAATCGGATCCATTTGAG 360

Qy 361 TTCTCGGATCGTGAGGAGAAAGCGCGCTTTCAGAGTGGCATGTTGGCGCGCGG 420
Db 361 TTCTCGGATCGTGAGGAGAAAGCGCGCTTTCAGAGTGGCATGTTGGCGCGCGG 420

Qy 421 GCATTGAGGCTTGGTCCGACAGGGAAGTGGACTCTTTGGTTGATGAGGATGAGTG 480
Db 421 GCATTGAGGCTTGGTCCGACAGGGAAGTGGACTCTTTGGTTGATGAGGATGAGTG 480

Qy 481 TCGTTAGGAGCGCGCTACTGCTTTCGCTTGAAGCGCGGGAATCGAGGCTGCAGCC 540
Db 481 TCGTTAGGAGCGCGCTACTGCTTTCGCTTGAAGCGCGGGAATCGAGGCTGCAGCC 540

Qy 541 CGCTCGGCGGAGCTCGTGCTTTTATCGAACTACACATTGAACAAGGACCGATCTC 600
Db 541 CGCTCGGCGGAGCTCGTGCTTTTATCGAACTACACATTGAACAAGGACCGATCTC 600

Qy 601 GAGCAGGAGCAATAGAGATCGGAGTGTAACTCCATCGTTGGCGTTTCGGCATTCGG 660
Db 601 GAGCAGGAGCAATAGAGATCGGAGTGTAACTCCATCGTTGGCGTTTCGGCATTCGG 660

Qy 661 GTTCCGCTCAAGGAGAGGAGCAAGCAGCGCGCAAAACCCCATGCTGCGGAGAT 720
Db 661 GTTCCGCTCAAGGAGAGGAGCAAGCAGCGCGCAAAACCCCATGCTGCGGAGAT 720

Qy 721 GCGTGTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCG 780
Db 721 GCGTGTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCG 780

Qy 781 GATGCAAGTGGCTACCGTTGGCCACTTCAAGTGGCGCGCGGTGGAGGCAACAGATC 840
Db 781 GATGCAAGTGGCTACCGTTGGCCACTTCAAGTGGCGCGCGGTGGAGGCAACAGATC 840

Qy 841 CCGGGGAGGTGGAATTACACTGACCTGCTGTTCTCGCATGAGGATCGCTCCGCTG 900
Db 841 CCGGGGAGGTGGAATTACACTGACCTGCTGTTCTCGCATGAGGATCGCTCCGCTG 900

Qy 901 CTGATCGACCGCATCTCGGTCATGCTGCGGAGGTGCGCTCCAGCGCGGTGTGCTGCC 960
Db 901 CTGATCGACCGCATCTCGGTCATGCTGCGGAGGTGCGCTCCAGCGCGGTGTGCTGCC 960

Db 61 TTCTCGGAGAGGCCCCGGGTGTTACCCGCGTGAACCTACACTCCAGAGCATGCCGCGG 120
QY 121 CGGGAACGCTCATTCGGCTATGAAAGCGCGCGCTTGAGCGCTTCGTGAAGACGCACTC 180
Db 121 CGGGAACGCTCATTCGGGCTATGAAAGCGCGCGCTTGAGCGCTTCGTGAAGACGCACTC 180
QY 181 GGAACATCATTCGGCGGAGCTGAAGGCACTGATCCGGAGCTTCCTCGCATCCGCGTGGT 240
Db 181 GGAACATCATTCGGCGGAGCTGAAGGCACTGATCCGGAGCTTCCTCGCATCCGCGTGGT 240
QY 241 TCACACTTCGATTCCTCGAAGCGCGGATGTTGATGGCACTCCAGCGGTGGTGC 300
Db 241 TCACACTTCGATTCCTCGAAGCGCGGATGTTGATGGCACTCCAGCGGTGGTGC 300
QY 301 GCCCTTGAGGCTGCCCGGCTGATGCTGGAGACGGCTACGTGAATCGGCATCCGATTTGAG 360
Db 301 GCCCTTGAGGCTGCCCGGCTGATGCTGGAGACGGCTACGTGAATCGGCATCCGATTTGAG 360
QY 361 TTCATCGCATCTGTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
Db 361 TTCATCGCATCTGTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
QY 421 GCCATTGCAAGGTTGTCGCCGACAGGAACCTGACTCTTTGTTGATGAGGATGAGTG 480
Db 421 GCCATTGCAAGGTTGTCGCCGACAGGAACCTGACTCTTTGTTGATGAGGATGAGTG 480
QY 481 TCCGTTAGGCGGCGCTACTGCTTCGGCTTGAAGCGCGGCAACTGCAGGCTGCAGCC 540
Db 481 TCCGTTAGGCGGCGCTACTGCTTCGGCTTGAAGCGCGGCAACTGCAGGCTGCAGCC 540
QY 541 CGCTCCGCGGCGACCTGCGTGTCTTTATCGAACTACATATGAACAGACCGATCCCTC 600
Db 541 CGCTCCGCGGCGACCTGCGTGTCTTTATCGAACTACATATGAACAGACCGATCCCTC 600
QY 601 GAGCAGGCAATAGAGATCGAGTTGTAACCTCATCGTTGGCGTTCGCGATTCGGG 660
Db 601 GAGCAGGCAATAGAGATCGAGTTGTAACCTCATCGTTGGCGTTCGCGATTCGGG 660
QY 661 GTTGCCGTCAAGGCGAGAGCGACACCGCGGCAACCCCATGACCTGCGCCAGGAT 720
Db 661 GTTGCCGTCAAGGCGAGAGCGACACCGCGGCAACCCCATGACCTGCGCCAGGAT 720
QY 721 GGCTGGTACCCGCGCTCTCATGTTGAGGAGGTCAACCGGTTCTGTCACAGATCGCC 780
Db 721 GGCTGGTACCCGCGCTCTCATGTTGAGGAGGTCAACCGGTTCTGTCACAGATCGCC 780
QY 781 GATGGCAGATGGCTACCGTTGGCCACCTCACAGTGGCGCCCGGTGGAGCAACCGGTC 840
Db 781 GATGGCAGATGGCTACCGTTGGCCACCTCACAGTGGCGCCCGGTGGAGCAACCGGTC 840
QY 841 CGGGGGAGGTGACCTTACACTGGACCTGCGTTCTCCGATCAGGAGTCCGCTCCCGGTG 900
Db 841 CGGGGGAGGTGACCTTACACTGGACCTGCGTTCTCCGATCAGGAGTCCGCTCCCGGTG 900
QY 901 CTGATCGACCGCATCTCGGTGATGGTTCGCGAGGTGCGCTCCAGGCGCGGTGGTGC 960
Db 901 CTGATCGACCGCATCTCGGTGATGGTTCGCGAGGTGCGCTCCAGGCGCGGTGGTGC 960
QY 961 GATGTGGATGAATTTTCAATCTCAGCGCGGTGCGAGTGGCTCTCCTACATGTGGAGCC 1020
Db 961 GATGTGGATGAATTTTCAATCTCAGCGCGGTGCGAGTGGCTCTCCTACATGTGGAGCC 1020
QY 1021 GTTTCGGAAGCGGCTCGGCTTGCAGTTTCACACCGGGATATCAGAGTGGGGCGGG 1080
Db 1021 GTTTCGGAAGCGGCTCGGCTTGCAGTTTCACACCGGGATATCAGAGTGGGGCGGG 1080
QY 1081 CAGACTCGATGTTTCATCGCCAGGTACGGAAGTGGTTCGTTCCGATCCGAGCCGT 1140
Db 1081 CAGACTCGATGTTTCATCGCCAGGTACGGAAGTGGTTCGTTCCGATCCGAGCCGT 1140
QY 1141 GCTGGCGGAGCAGCTTCCGGAAGATGAGCCGATTTGATGACCTTCGCAAGGAACT 1200
Db 1141 GCTGGCGGAGCAGCTTCCGGAAGATGAGCCGATTTGATGACCTTCGCAAGGAACT 1200

QY 1201 GAGGTTGTCTCTCGGTAATGAAGCACTTGACCGGTAA 1239
Db 1201 GAGGTTGTCTCTCGGTAATGAAGCACTTGACCGGTAA 1239
RESULT 3
BD194451 1239 bp DNA linear PAT 17-JUL-2003
LOCUS Arthrobacter aureusens-origin recombinant L-N-carbamoylase and
DEFINITION process for producing L-amino acid thereby.
ACCESSION BD194451
VERSION JP 2002510484-A/4.
KEYWORDS Arthrobacter aureusens
SOURCE Arthrobacter aureusens
ORGANISM Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococccineae; Micrococccaceae; Arthrobacter.
REFERENCE 1 (bases 1 to 1239)
AUTHORS Artenboofner, J., Matteu, R., Beash, M., Jillduck, C., Vize, A. and
Willms, B.
TITLE Arthrobacter aureusens-origin recombinant L-N-carbamoylase and
JOURNAL process for producing L-amino acid thereby
PATENT: JP 2002510484-A 4 09-APR-2002;
DEGUSA HUELS AG, UNIVERSITÄT STUTTGART, ROCHE DIAGNOSTICS GMBH
COMMENT OS Arthrobacter aureusens
PN JP 2002510484-A/4
PD 09-APR-2002
PF 11-MAR-1999 JP 2000542435
PR 02-APR-1998 DE 198 14 813.5
PI JOSEPH ARTENBOOFNER, RALPH MATTEU, MARCUS BEASH, CHRISTOPHE PI
JILLDUCK,
PI ARNYA VIZE, BURCART WILLMS
PC C12N15/09, C12N1/21, C12N9/80, C12P13/04, C12P41/00, C12N11/10, PC
(C12N9/80, C12R1:06), C12N15/00
CC Arthrobacter aureusens-origin recombinant
L-N-carbamoylase and
CC process for
CC producing L-amino acid thereby
FH Key Location/Qualifiers
FT CDS (1)..(1239).
FEATURES
source
1..1239
/organism="Arthrobacter aureusens"
/mol_type="genomic DNA"
/db_xref="taxon:43663"
ORIGIN
Query Match 100.0%; Score 1239; DB 6; Length 1239;
Best Local Similarity 100.0%; Pred. No. 8.9e-208;
Matches 1239; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GTGACCTCTCAGAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCG 60
Db 1 GTGACCTCTCAGAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCG 60
QY 61 TTCTCGGAGAGAGCGCGCTGTACCGGCTGACCTACACTCCAGAGCATCCGCGG 120
Db 61 TTCTCGGAGAGAGCGCGCTGTACCGGCTGACCTACACTCCAGAGCATCCGCGG 120
QY 121 CGGGAACGCTCATTCGGCTATGAAAGCGCGCTTCGAGCGTTCGTGAAGACGCACTC 180
Db 121 CGGGAACGCTCATTCGGCTATGAAAGCGCGCTTCGAGCGTTCGTGAAGACGCACTC 180
QY 181 GGAACATCATTCGGCGGAGCTGAAGGCACTGATCCGGAGCTTCCTCGCATCCGCGTGGT 240
Db 181 GGAACATCATTCGGCGGAGCTGAAGGCACTGATCCGGAGCTTCCTCGCATCCGCGTGGT 240
QY 241 TCACACTTCGATTCCTCGAAGCGCGGATGTTTGTGGCACTCCAGCGGTGGTGC 300
Db 241 TCACACTTCGATTCCTCGAAGCGCGGATGTTTGTGGCACTCCAGCGGTGGTGC 300
QY 301 GCGCTTGAGGCTGCCCGGCTGATGCTGGAGACGGCTACGTGAATCGGCATCCGATTTGAG 360

Db 301 GGCCTTGAGCTGCCCGGCTGATCTGGAGAACGGCTACGTGAATCGGCATCCATTGAG 360
Qy 361 TTTCATCGCATCGTGGAGAGAAAGGGCCCGCTTTCAGCAGTGGCATGTTGGCGCGCGG 420
Db 361 TTTCATCGCATCGTGGAGAGAAAGGGCCCGCTTTCAGCAGTGGCATGTTGGCGCGCGG 420
Qy 421 GCATTGCAAGGTTGGTCCGACAGGGAACCTGGACTCTTTGGTTGATGAGGATGAGTG 480
Db 421 GCATTGCAAGGTTGGTCCGACAGGGAACCTGGACTCTTTGGTTGATGAGGATGAGTG 480
Qy 481 TCCGTTAGGACGCGCTACTCGCTTCGCTTGAAGCGGGCGAACTGCAGGCTCGAGCC 540
Db 481 TCCGTTAGGACGCGCTACTCGCTTGAAGCGGGCGAACTGCAGGCTCGAGCC 540
Qy 541 CGCTCCGCGCGGACCTCGCTGCTTTTATCGAACTACACATTTGAACAAAGGACCGATCCTC 600
Db 541 CGCTCCGCGCGGACCTCGCTGCTTTTATCGAACTACACATTTGAACAAAGGACCGATCCTC 600
Qy 601 GAGCAGGACCAATAGAGATCGAGTTGTAACTTCATCGTTGGCGTTCCGCGATTCGGG 660
Db 601 GAGCAGGACCAATAGAGATCGAGTTGTAACTTCATCGTTGGCGTTTCGCGCATTTGGCGG 660
Qy 661 GTTTCGCTCAAAGGACAGGACACACCGCGGCACACCCCATGCACTGCGCCAGGAT 720
Db 661 GTTTCGCTCAAAGGACAGGACACACCGCGGCACACCCCATGCACTGCGCCAGGAT 720
Qy 721 GCGCTGGTACCCGCGCTCTCATGTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCC 780
Db 721 GCGCTGGTACCCGCGCTCTCATGTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCC 780
Qy 781 GATGGACAGTGGCTACGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACAGGTC 840
Db 781 GATGGACAGTGGCTACGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACAGGTC 840
Qy 841 CCGGGGAGGTGACCTTCACTGACCTCGCTTCTCGCATGAGGAGTCCGTCGCGGTG 900
Db 841 CCGGGGAGGTGACCTTCACTGACCTCGCTTCTCGCATGAGGAGTCCGTCGCGGTG 900
Qy 901 CTGATCGACCGCATCTCGCTCATGTGCGCGAGGTCCGCTCCAGGCGGTTGCTGCC 960
Db 901 CTGATCGACCGCATCTCGCTCATGTGCGCGAGGTCCGCTCCAGGCGGTTGCTGCC 960
Qy 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGCTGGCTTCCATCATGTTGAGCGCC 1020
Db 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGCTGGCTTCCATCATGTTGAGCGCC 1020
Qy 1021 GTTTCGGAAGCGGCTCGGCTTGCAGTTTCACACACCGGATATCAGCAGTGGGCGGCG 1080
Db 1021 GTTTCGGAAGCGGCTCGGCTTGCAGTTTCACACACCGGATATCAGCAGTGGGCGGCG 1080
Qy 1081 CACGACTCGATTTTCATGCCCGAGGTACGAGCTGGTTCGTTTCGTTCCAAAGCGGT 1140
Db 1081 CACGACTCGATTTTCATGCCCGAGGTACGAGCTGGTTCGTTTCGTTCCAAAGCGGT 1140
Qy 1141 GCTGCGCGAGCACGTTCCCAAGATGAGCCGATTCGATGACTTCGCAAGGAACT 1200
Db 1141 GCTGCGCGAGCACGTTCCCAAGATGAGCCGATTCGATGACTTCGCAAGGAACT 1200
Qy 1201 GAGGTTGCTCCTCGGGTAATGAGGCACCTTGACCGGTAA 1239
Db 1201 GAGGTTGCTCCTCGGGTAATGAGGCACCTTGACCGGTAA 1239

RESULT 4
AX103678
LOCUS AX103678 1239 bp DNA linear PAT 30-APR-2001
DEFINITION Sequence 7 from Patent WO0123582.
ACCESSION AX103678
VERSION AX103678.1 GI:13919890
KEYWORDS .
SOURCE Arthrobacter aureus
ORGANISM Arthrobacter aureus

Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococcales; Micrococcaceae; Arthrobacter.
1 Altenbuchner,J., Mattes,R., Syltatk,C., Wiese,A., Wilms,B.,
Bommarius,A. and Tischer,W.
Whole cell catalyst comprising a hydantoinease, a racemase and a
carbamoylase
Patent: WO 0123582-A 7 05-APR-2001;
Degussa (DE) ; Universitaet Stuttgart (DE) ; Roche Diagnostics GmbH
(DE)
Location/Qualifiers
1. .1239
/organism="Arthrobacter aureus"
/mol_type="unassigned DNA"
/db_xref="taxon:43663"
1. .1239
/note="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC37705.1"
/db_xref="GI:13919891"
/translation="MTLQKAQARIEKIEIRLSRFSAGPGVTRLTTPPEHAAARETL
IAAMKAALSVDNALGNIIGRREGTDPFLPAIAGVGHFSDVRNGMDFGTAGVVCAL
EAAVMLENGVNRHPFEITAVEEGARSSGMLGGRAIAGLVADRELDVLVDGV
SVRAQATAFGLKPELQAAARSADIRAFIELHTEQPILEQEOIEIGVVTSLVGRV
LRVAKGRSDHAGTTPMHLRQDALVPAALMVRNRFNEIADGTVAITVGHLTVPAGG
GNQVPGSVDFTLDRSPHELSRLVLDRIISMVGEVASQAGVADDEFFNLSPVQLA
PTWDAVREASALQFTHRDISSGAGHDSMFIAQVTDVMGVFVPSRAGRSHVPEWTD
FDDLKGTKEVTVLRVVKALDR"
ORIGIN
Query Match 99.9%; Score 1238; DB 6; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.3e-207;
Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 2 TGACCCCTGCAGAAAGCGCAGCGCGCATTTGAGAAGAGATCCGGAGGCTCTCCCGGT 61
Db 2 TGACCCCTGCAGAAAGCGCAGCGCGCATTTGAGAAGAGATCCGGAGGCTCTCCCGGT 61
Qy 62 TCTCGCAGAAAGCGCGCGGTGTATCCCGGTGACCTACACTCCAGAGCATCCCGCGCGC 121
Db 62 TCTCGCAGAAAGCGCGCGGTGTATCCCGGTGACCTACACTCCAGAGCATCCCGCGCGC 121
Qy 122 GGGAAAACGCTCATTTGGGCTATGAAGCGCGCGCTTGGAGGTTCTGTAAGACGCACTCG 181
Db 122 GGGAAAACGCTCATTTGGGCTATGAAGCGCGCGCTTGGAGGTTCTGTAAGACGCACTCG 181
Qy 182 GAAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTGCGATCCGGTCGGTT 241
Db 182 GAAACATCATCGCGCGAGCTGAAGGCACTGATCCGAGCTTCTCTGCGATCCGGTCGGTT 241
Qy 242 CACACTTCGATTTCTGTCGAAACCGCGGGATGTTTGTATGGCACTGCAAGCGTGTGTGCG 301
Db 242 CACACTTCGATTTCTGTCGAAACCGCGGGATGTTTGTATGGCACTGCAAGCGTGTGTGCG 301
Qy 302 CCCTTGAGGCTGCCCGGCTGATGCTGGAGAACGCGCTACGTAATCGGCATCCATTTGAGT 361
Db 302 CCCTTGAGGCTGCCCGGCTGATGCTGGAGAACGCGCTACGTAATCGGCATCCATTTGAGT 361
Qy 362 TCATCGCGATCGTGGAGGAGGAGCGCGCGCTTTCAGCAGTGGCATGTTGGCGCGCGGG 421
Db 362 TCATCGCGATCGTGGAGGAGGAGCGCGCGCTTTCAGCAGTGGCATGTTGGCGCGCGGG 421
Qy 422 CCATTGAGGGTGGTTCGCCGACAGGGAACCTGGAATCTTTGGTTGATGAGATGGAGTGT 481
Db 422 CCATTGAGGGTGGTTCGCCGACAGGGAACCTGGAATCTTTGGTTGATGAGATGGAGTGT 481
Qy 482 CCGTTAGGCGCGGCTACTGCTTCGGCTTGAAGCGCGGCGAACTGCGAGCTCGAGCCC 541
Db 482 CCGTTAGGCGCGGCTACTGCTTCGGCTTGAAGCGCGGCGAACTGCGAGCTCGAGCCC 541
Qy 542 GCTCCGCGCGGACCTGCGTGTCTTTTATCGAACTACACATTTGAACAAAGGACCGATCCTCG 601

Db 542 GCTCCCGCGGACCTGCGTGTCTTTATCGAACTACACATTAAGACCGATCCTCG 601
QY 602 AGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCTGTTGCGCTTCGCGCATTCGCGG 661
Db 602 AGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCTGTTGCGCTTCGCGCATTCGCGG 661
QY 662 TTGCGCTCAAGAGCAGAAGCAGCACGCGCGGCAACACCCCATGCACTTGGCCAGGATG 721
Db 662 TTGCGCTCAAGAGCAGAAGCAGCACGCGCGGCAACACCCCATGCACTTGGCCAGGATG 721
QY 722 CGCTGTACCCCGCTCTCATGTTGAGGAGTCAACCGGTTCTGTTCAACGAGATCGCGG 781
Db 722 CGCTGTACCCCGCTCTCATGTTGAGGAGTCAACCGGTTCTGTTCAACGAGATCGCGG 781
QY 782 ATGGCACAGTGGCTACCGTTGGCCACTCACAGTGGCCCGCTGGAGGCAACAGGTCC 841
Db 782 ATGGCACAGTGGCTACCGTTGGCCACTCACAGTGGCCCGCTGGAGGCAACAGGTCC 841
QY 842 CGGGGAGGTGACTTTCACATGAGCTCGGTTCTCCGATGAGGAGTCCGCTCCGCGTGC 901
Db 842 CGGGGAGGTGACTTTCACATGAGCTCGGTTCTCCGATGAGGAGTCCGCTCCGCGTGC 901
QY 902 TGATCGACGCATCTCGGTATGTCGGGAGTCCCTCCAGGCGGCTGGCTGCGG 961
Db 902 TGATCGACGCATCTCGGTATGTCGGGAGTCCCTCCAGGCGGCTGGCTGCGG 961
QY 962 ATGTGATGAATTTTCAATCTCAGCCCGTGGCTGCTCTACCATGTTGGAGCGCG 1021
Db 962 ATGTGATGAATTTTCAATCTCAGCCCGTGGCTGCTCTACCATGTTGGAGCGCG 1021
QY 1022 TTTCGGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGCAGTGGGGCGGCG 1081
Db 1022 TTTCGGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGCAGTGGGGCGGCG 1081
QY 1082 AGCACTCGATGTTTCATCGCCAGGTTCAGGAGTTCAGGAGTGGTTTCGTTCAAGCCGTG 1141
Db 1082 AGCACTCGATGTTTCATCGCCAGGTTCAGGAGTTCAGGAGTGGTTTCGTTCAAGCCGTG 1141
QY 1142 CTGGCGGAGCAGCTTCGCGGAGATGACCGATTTCCATGACCTTCGCAAGGAGT 1201
Db 1142 CTGGCGGAGCAGCTTCGCGGAGATGACCGATTTCCATGACCTTCGCAAGGAGT 1201
QY 1202 AGGTTGTCTCCGGGTAATGAAGGACTTGAACCGGTA 1239
Db 1202 AGGTTGTCTCCGGGTAATGAAGGACTTGAACCGGTA 1239

RESULT 5
AX370705
LOCUS AX370705 1239 bp DNA linear PAT 01-MAR-2002
DEFINITION Sequence 3 from Patent WO0210424.
ACCESSION AX370705
VERSION AX370705.1 GI:19168862
KEYWORDS
SOURCE Arthrobacter aureus
ORGANISM Arthrobacter aureus
Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococcales; Micrococcaceae; Arthrobacter.
REFERENCE
1
AUTHORS Krimmer, H.P., May, O., Klement, I., Drauz, K. and Reichert, D.
TITLE Process for the preparation of allylsine acetal
JOURNAL Patent: WO 0210424-A 3 07-FEB-2002;
Degussa AG (DE)
FEATURES
source
1. .1239
/organism="Arthrobacter aureus"
/mol_type="unassigned DNA"
/db_xref="taxon:43663"
ORIGIN
Query Match 99.9%; Score 1238; DB 6; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.3e-207;

Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TGACCCCTGCAGAAAGCGCAAGCGCGCATTTAGAGAAAGAGATCCGGAGCTCTCCCGGT 61
Db 2 TGACCCCTGCAGAAAGCGCAAGCGCGCATTTAGAGAAAGAGATCCGGAGCTCTCCCGGT 61
QY 62 TCTCGCAGAAAGCGCCCGGTGTTTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGCG 121
Db 62 TCTCGCAGAAAGCGCCCGGTGTTTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGCG 121
QY 122 GGGAAACGCTCATTTGGCGGTATGAAGCGCGCCCTTTGAGCGTTCTGTGAAGACGACTCG 181
Db 122 GGGAAACGCTCATTTGGCGGTATGAAGCGCGCCCTTTGAGCGTTCTGTGAAGACGACTCG 181
QY 182 GAAACATCATTCGCGGACGCTGAAGGACGCTGATCGGAGCTTCTCGGATCGCGGTGCGTT 241
Db 182 GAAACATCATTCGCGGACGCTGAAGGACGCTGATCGGAGCTTCTCGGATCGCGGTGCGTT 241
QY 242 CACACTTCGATTTCTGTCCGAAACGCGCGGATGTTTATGGCACTGCAGGCGTGGTGTGG 301
Db 242 CACACTTCGATTTCTGTCCGAAACGCGCGGATGTTTATGGCACTGCAGGCGTGGTGTGG 301
QY 302 CCTTTGAGCTCGCCGGGTGATCTGGAGAAAGGCTACGTAATCGGCATTCATTTGAGT 361
Db 302 CCTTTGAGCTCGCCGGGTGATCTGGAGAAAGGCTACGTAATCGGCATTCATTTGAGT 361
QY 362 TCATCGCATCTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGCGG 421
Db 362 TCATCGCATCTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGCGG 421
QY 422 CCAATTCAGGCTTGGTTCGCGCAGGAACTGAGACTCTTTTGGTTGATGAGGATGAGTGT 481
Db 422 CCAATTCAGGCTTGGTTCGCGCAGGAACTGAGACTCTTTTGGTTGATGAGGATGAGTGT 481
QY 482 CGTTAGGAGCAGCGGCTACTGCTTCGCTTGAAGCGCGGCGAACTGAGGCTGACGCC 541
Db 482 CGTTAGGAGCAGCGGCTACTGCTTCGCTTGAAGCGCGGCGAACTGAGGCTGACGCC 541
QY 542 GCTCCGCGCGGACCTGCGTGTCTTTATCGAACTACACATTTGAAAGGACCGATCCTCG 601
Db 542 GCTCCGCGCGGACCTGCGTGTCTTTATCGAACTACACATTTGAAAGGACCGATCCTCG 601
QY 602 AGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCTGTTGCGCTTCGCGCATTCGCGG 661
Db 602 AGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCTGTTGCGCTTCGCGCATTCGCGG 661
QY 662 TTGCGCTCAAGAGCAGAAGCAGCACGCGCGGCAACACCCCATGCACTTGGCCAGGATG 721
Db 662 TTGCGCTCAAGAGCAGAAGCAGCACGCGCGGCAACACCCCATGCACTTGGCCAGGATG 721
QY 722 CGCTGTACCCCGCTCTCATGTTGAGGAGTCAACCGGTTCTGTTCAACGAGATCGCGG 781
Db 722 CGCTGTACCCCGCTCTCATGTTGAGGAGTCAACCGGTTCTGTTCAACGAGATCGCGG 781
QY 782 ATGGCACAGTGGCTACCGTTGGCCACTCACAGTGGCCCGCTGGAGGCAACAGGTCC 841
Db 782 ATGGCACAGTGGCTACCGTTGGCCACTCACAGTGGCCCGCTGGAGGCAACAGGTCC 841
QY 842 CGGGGAGGTGACTTTCACATGAGCTCGGTTCTCCGATGAGGAGTCCGCTCCGCGTGC 901
Db 842 CGGGGAGGTGACTTTCACATGAGCTCGGTTCTCCGATGAGGAGTCCGCTCCGCGTGC 901
QY 902 TGATCGACGCATCTCGGTATGTCGGGAGTCCCTCCAGGCGGCTGGCTGCGG 961
Db 902 TGATCGACGCATCTCGGTATGTCGGGAGTCCCTCCAGGCGGCTGGCTGCGG 961
QY 962 ATGTGATGAATTTTCAATCTCAGCCCGTGGCTGCTCTACCATGTTGGAGCGCG 1021
Db 962 ATGTGATGAATTTTCAATCTCAGCCCGTGGCTGCTCTACCATGTTGGAGCGCG 1021
QY 1022 TTTCGGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGCAGTGGGGCGGCG 1081
Db 1022 TTTCGGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGCAGTGGGGCGGCG 1081

Query Match Best Local Similarity Matches 1238; Conservative 0; Mismatches 1; Indels 0; Gaps 0;	QY	1	GTGACCTCGCAGAAAGCGGCGCGCATTTGAGAAAGAGATCCGGGAGCTCTCCCGG 60	Db	6024	GTGACCTCGCAGAAAGCGGCGGAGCGCATTTGAGAAAGAGATCCGGGAGCTCTCCCGG 6083
QY	Db	61	TTCCTGGCAGAGGCGCGCGTGTACCGGCTGACCTACACTCCAGAGCATGCCCGCGG 120	Db	6084	TTCCTGGCAGAGGCGCGCGTGTACCGGCTGACCTACACTCCAGAGCATGCCCGCGG 6143
QY	Db	121	CGGGAAACGCTATTCGGGCTATGAAGCGCGCGCTTGAGCGTTGCGTGAAGACGCATC 180	Db	6144	CGGGAAACGCTATTCGGGCTATGAAGCGCGCGCTTGAGCGTTGCGTGAAGACGCATC 6203
QY	Db	181	GGAAACATCATCGGCGCAGCTGAAGCAGCTGATCCGGAGCTTCTCGGATCGGGTCGGT 240	Db	6204	GGAAACATCATCGGCGCAGCTGAAGCAGCTGATCCGGAGCTTCTCGGATCGGGTCGGT 6263
QY	Db	241	TCACATTCGATTTCTGTCGAAACCGCGGGATGTTTGTGGCACTCAGCGCGTGTGTC 300	Db	6264	TCACATTCGATTTCTGTCGAAACCGCGGGATGTTTGTGGCACTCAGCGCGTGTGTC 6323
QY	Db	301	GCCCTTGAGCGTCCCGGCTGATGCTGGAGACGGCTACGTGAATCGGCATCCATTGAG 360	Db	6324	GCCCTTGAGCGTCCCGGCTGATGCTGGAGACGGCTACGTGAATCGGCATCCATTGAG 6383
QY	Db	361	TTCATCGCATCTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420	Db	6384	TTCATCGCATCTGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGG 6443
QY	Db	421	GCCATTGCAGGGTTGTCGCCGACAGGGAACTCGGACTCTTTTGGTTGATGAGGATGAGTG 480	Db	421	GCCATTGCAGGGTTGTCGCCGACAGGGAACTCGGACTCTTTTGGTTGATGAGGATGAGTG 480

VSDGELFEIPEIQAACGSVIVVAENETIIQALQKQIKAAAGGKMDMAAYEASQVPOFN
EATQALLQKEACGLIHLVNSNPDGVELIHOAQSEGQDVHCESPQVLIINTIDDAE
RIGYMKVAPVRASAKNVLQENGLVITDLSHGGHPVEDKEPGWKDVMKAGNG
ALGLETSLPMLTNGNKGRLSLERLVEVMCEKPAKLFGLYPOKGTLOWGSDADLIL
DLIDITKVDASQFRLHKYSPFDGMPVTGAPVLITWVRGTVVAEQEGVQGFVT
RHHYEASK"
6024..7262
/gene="hyuC"
6024..7262
/gene="hyuC"
/codon_start=1
/transl_table=11
/product="L-N-carbamoylase HyuC"
/protein_id="AA02131.1"
/db_xref="GI:9931314"
/translations="MTLQKAQAEIKELRSFSAEGFVGVTRLTYTPEHAARETL
IAAMKAALSVDALNIIIRREGIDPELPAVAGSHFSDVNRGMPDGTAGVVAL
EAAVLENGYVNRHPEPIAIVEEGARPSSGMLGRAIAGLVADRELDLSLDEGV
SVQAATAGLKEGELQAAARSANDURATFELHIEQPILEQIEIGVVTSLVGRG
LRVAVKGRSDHAGTTPMHLQRQDALVPAALMVRVNRVNEIADGTVAIVGHLTVAPGG
GNQVGEVDFTLDRSPHEESLRVLIDRISVMVGEVASQAGVAADVEFNLSPVOLA
PTMVDVRAASALQFTHRIDISSGAGHDSMFIQAVTDVGMVFPVPSRAGRSHVPEEWD
FDLRKGTETVLAVMKALDR"
complement(7416..8450)
/notes="hypothetical protein"
/codon_start=1
/evidence=not experimental
/transl_table=11
/product="laci repressor homolog"
/protein_id="AA02132.1"
/db_xref="GI:9931315"
/translations="MMTEQLKRAATISDVAEHPGVRAAVSKVLNAYGVSPAMRTKV
ESMTLLNTPRALARGMRNSYTLGVFLVDSNTFFSLAEGIRDVAELGQVVLIG
QARQGLDAQOKLT EAMVDRMDGLVLIAPFPNDELEKLGRTIPVVVLGRHGPXKFPD
TVASDDIAGSALIVDLHLSLGHIERISMLANTOPDSEPGMEOTVRRHGYDOAMEKHGLA
KYIDHVPGRWGTESGLEAGRSLSAARPLAQRPSALHGGADVAALGATYALFEAGIHIPD
DMSIVGDNIPATGLPPIGLSTVDQSGVMGROAELLISRIEGRTESESHVLVSPRLI
VRKTSAPRR"

ORIGIN

Query Match 99.9%; Score 1237.4; DB 1; Length 8715;
Best Local Similarity 99.9%; Pred. No. 1.3e-207;

Db	6444	GCCATTGCAGGGTTGGTCGCCGACAGGGAAC	TTCGCT
----	------	---------------------------------	--

source		1. .1263
		/organism="Arthrobacter aureusens"
		/mol_type="unassigned DNA"
		/db_xref="taxon:43663"
ORIGIN		
Query Match		96.2%; Score 1192.4; DB 6; Length 1263;
Best Local Similarity		97.8%; Pred. No. 1.4e-199;
Matches 1208; Conservative		0; Mismatches 27; Indels 0; Gaps 0;
QY	2	TGACCTCGAGAAAGCGGCGCGCATTTGAGAAAGATTCGGGAGCTCTCCGGT 61
Db	2	TGACCTCGAGAAAGCGGCGCGCATTTGAGAAAGATTCGGGAGCTCTCCGGT 61
QY	62	TCTCGCAGAGCCCGGTGTACCCGGCTACCTACACTCCAGAGCATCGCGCGCG 121
Db	62	TCTCGCAGAGCCCGGTGTACCCGGCTACCTACACTCCAGAGCATCGCGCGCG 121
QY	122	GGGAAACGCTCATTTGGCGCTATGAAAGCGCGCGCTTGGAGCGTTCTGAAAGCGCTCTCG 181
Db	122	GGGAAACGCTCATTTGGCGCTATGAAAGCGCGCGCTTGGAGCGTTCTGAAAGCGCTCTCG 181
QY	182	GAAACATCATCGCGCGCATGTAAGCACTGATCGGAGCTTCTCGGATCGCGTGGTT 241
Db	182	GAAACATCATCGCGCGCATGTAAGCACTGATCGGAGCTTCTCGGATCGCGTGGTT 241
QY	242	CACACTTCGATCTGTCGAAACGGCGGATGTTGATGGCACTGCAGCGTGGTGGCG 301
Db	242	CACACTTCGATCTGTCGAAACGGCGGATGTTGATGGCACTGCAGCGTGGTGGCG 301
QY	302	CCCTTGAGGCTCCCGGCTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTGGAT 361
Db	302	CCCTTGAGGCTCCCGGCTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTGGAT 361
QY	362	TGATCGCGATCTGGAGAGAGAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
Db	362	TGATCGCGATCTGGAGAGAGAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
QY	422	CATTTCAGGTTGTCGCGACAGGAACTGCACTCTTTGCTGATGAGGATGAGTGT 481
Db	422	CATTTCAGGTTGTCGCGACAGGAACTGCACTCTTTGCTGATGAGGATGAGTGT 481
QY	482	CCGTTAGGCGGCGCTACTGCTTTCGGCTTGAAGCGCGCGCAACTGCAGGCTGCAGGCC 541
Db	482	CCGTTAGGCGGCGCTACTGCTTTCGGCTTGAAGCGCGCGCAACTGCAGGCTGCAGGCC 541
QY	542	GCTCCGCGCGACCTCGCTGCTTTATCGAACTACATTGAAACAGGACCGATCCTCG 601
Db	542	GCTCCGCGCGACCTCGCTGCTTTATCGAACTACATTGAAACAGGACCGATCCTCG 601
QY	602	AGCAGAGCAATAGAGATCGGAGTTGTAACCTCCATCTTCGCGCTTCGCGATTGCGGG 661
Db	602	AGCAGAGCAATAGAGATCGGAGTTGTAACCTCCATCTTCGCGCTTCGCGATTGCGGG 661
QY	662	TTGCCGCTCAAAGCGAGACGACACCGCGGCACACCCCATGCACCTGCGCCAGGATG 721
Db	662	TTGCTGCTCAAAGCGAGACGACACCGCGGCACACCCCATGCACCTGCGCCAGGATG 721
QY	722	CGCTGGTACCCCGCTCTCATGTTGAGGAGGATCAACCGGTTCTCAACGAGATCGCG 781
Db	722	CGCTGGTACCCCGCTCTCATGTTGAGGAGGATCAACCGGTTCTCAACGAGATCGCG 781
QY	782	ATGGCAGATGCTACCGTTGGCCACCTACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
Db	782	ATGGCAGATGCTACCGTTGGCCACCTACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
QY	842	CGGGGAGGTGGAATTCACATCGGACCTGCTTCGCGCATGAGGATCGCTCCCGGTG 901
Db	842	CGGGGAGGTGGAATTCACATCGGACCTGCTTCGCGCATGAGGATCGCTCCCGGTG 901
QY	902	TGATCGACCGCATCTCGGTTCGGCGAGGTGCGCTCCAGCGCGGTGTGGTGGCG 961
Db	902	TGATCAACCGCATCTCGGTTCGGCGAGGTGCGCTCCAGCGCGGTGTGGTGGCG 961

QY	962	ATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGTGGCTCTTACCAATGCTGACGCGG 1021
Db	962	ATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGTGGCTCTTACCAATGCTGACGCGG 1021
QY	1022	TTCCGGAAGCGGCTCGGCTTGCAGTTTCACACACCGGATATCAGCAGTGGGCGGCGC 1081
Db	1022	TTCCGGAAGCGGCTCGGCTTGCAGTTTCACACACCGGATATCAGCAGTGGGCGGCGC 1081
QY	1082	ACGACTCGATTTTCATCGCCCGGCTCAGGACGTCGGATGGTTTTCGTTCCAAAGCGGTG 1141
Db	1082	ACGACTCGATTTTCATCGCCCGGCTCAGGACGTCGGATGGTTTTCGTTCCAAAGCGGTG 1141
QY	1142	CTGGCCGAGGACGCTTCCGGAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACTG 1201
Db	1142	CTGGCCGAGGACGCTTCCGGAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACTG 1201
QY	1202	AGTTGTCTCTCCGGTAAATGAAGCACTTGACCGG 1236
Db	1202	AGTTGTCTCTCCGGTAAATGAAGCACTTGACCGG 1236
RESULT 8		
AX397881		
LOCUS		AX397881 1239 bp DNA linear PAT 27-MAY-2002
DEFINITION		Sequence 5 from Patent EP1188826.
ACCESSION		AX397881
VERSION		AX397881.1 GI:21260748
KEYWORDS		Microbacterium liquefaciens
SOURCE		Microbacterium liquefaciens
ORGANISM		Microbacterium liquefaciens
REFERENCE		1 Suzuki, S.C., Onishi, N.C. and Yokozeki, K.C.
AUTHORS		5-substituted hydantoin racemase, dna coding for the same, and
TITLE		process for producing optically active amino acids
JOURNAL		Patent: EP 1188826-A 5 20-MAR-2002;
FEATURES		Ajinomoto Co., Inc. (JP)
source		Location/Qualifiers
		1. .1239
		/organism="Microbacterium liquefaciens"
		/mol_type="unassigned DNA"
		/db_xref="taxon:33918"
CDS		1. .1239
		/note="unnamed protein product; Chase"
		/codon_start=1
		/transl_table=11
		/protein_id="CAD32595.1"
		/db_xref="GI:21260749"
		/db_xref="REMTREMBL:CAD32595"
		/translation="MTLQARADRIESELWTLRFRFVSGPQVTRLTYPPEHAAAREVI
		VAAMQRTGLSVHEDALGNIIIGRRGSDPALPAIFGSHFDSVRNGMDFDGTAGVVCAL
		EAARVLQESGYVNRHLEIVAEVEGFRFSSGMLGRLAIGLVSDADLDTLVDEGVL
		TRREATAFGLEPEGLTAARDDRLAFIELHIEQGFILFEQKEVEIGVVTGIVGVA
		FRIIVEGRSDHAGTTPMHLRQDALVPAALMVRINRNVNEIADCTVATVGLHTVPGG
		LNQVPGVVEFTDLRSPEEIRLLVDRIEAMVAEVAAGVAAVNGFFALSPVGLS
		PVVVDVRDRAASELGFTHRDITSGAHDSEMFIAQITDGMVVFPSRAGRSHVPEMSD
		FDDLKRGTDVVLHVVTALDR"
ORIGIN		
Query Match		59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity		74.8%; Pred. No. 4e-120;
Matches 927; Conservative		0; Mismatches 312; Indels 0; Gaps 0;
QY	1	GTGACCTTCAGAAAGCGGCGCGCATTCAGAAAGATTCGGGAGCTCTCCCGG 60
Db	1	GTGACCTTCAGAGGCGCGCGCATTCGATCGAGGAGGAGCTCTGGAATCTCTCCCGC 60
QY	61	TTCTCGCAGAGGCGCGGTGTTTACCGGCTGACCTACACTCCAGAGCATGCCCGCG 120
Db	61	TTCTCGGTCGAAAGGCGCGGCTGACACGCTCTACGTACTCCGGAGCAGCGCGCG 120

QY 121 CGGAAACCGCTCATTCGGCGTATGAAGCGCGCGCTTGAGCGCTTCGTGAAGACGCACATC 180
DB 121 CGAGAGGTGATCGTCGCCGCTATGACGCGACCGGGCTGAGCGCTCCACGAGGACGCTCTC 180
QY 181 GGAACATCATCGGCGAGTGAAGCACTGATCCGGAGCTTCCTCGGATCGCGGTCCGT 240
DB 181 GGCACATCATCGTCCGGTGAAGGAGGAGCCCGCTCTGCGGCGATCGCTTCGGC 240
QY 241 TCACACTTCGATTCCTCCGAAACCGCGGATGTTTGTGAGCACTGACGCGGTGTGTGC 300
DB 241 TCGCACTTCGACTCGCTCGCAACCGCGGATGTTTCGACGGCACCGCGGCGTGTGTGC 300
QY 301 GCCCTTAGCGCTCCCGGCTGATCTGGAACCGCTACGTGAATCGGATCCATTTAG 360
DB 301 CGCTCGAGGCTCGAGGGTGTGACGAGAGCGGATATGTGAACCGTATCTCTTCGAG 360
QY 361 TTCATCGCATCGTGCAGAGGAGGCGCGCTTCAGAGTGGCATGTTGGCGCGCGG 420
DB 361 GTCATCGCATGCTGAAGAGGAGGACCGCTTCAGCAGCGCATGCTGGCGGTCCG 420
QY 421 GCCATTGACGGTGTGTCGCCGACAGGAACTGGAATCTTTGTTGATGAGATGAGTG 480
DB 421 GCGATCGCGGGCTCGTGTCCGACCGCATCTGACACCCCTGTTGACGACGACGCGTG 480
QY 481 TCGCTTAGCAGCGGCTACTGCTTCGGCTTGAAGCGCGGCGAACTGCAAGGTGCGAGCC 540
DB 481 ACGGTGCGGAGCGGCCACGGCTTCGGGCTGGAACCGGCTGAGCTGCGGACGGCGGC 540
QY 541 CGCTCGCGCGGACCTGGTGTCTTATCGAACTACATACATGAACGAGCGATCTC 600
DB 541 CGTACGAGGATGACCTTCGGCTTCATCGAGTGTGCATCGAGCGGGCGGATCTCTC 600
QY 601 GAGCAGGACAAATAGAGATCGAGTGTAACTCCATCTGTTGGCGCTTCGCGATTCGG 660
DB 601 GAGCAGGAGAGTGGAGATCGGCTCGTACGGGGATGCTCGGTTCGCGCTTCGG 660
QY 661 GTTGGCGTCAAGGCAAGAGACCAACGCGGCAACCCCGCATGCACTTCGCGCCAGGAT 720
DB 661 ATCACGTTGAGGCGGAGGACCAACGCGGACGACCCCATGCACTTCGCGGACGAG 720
QY 721 GCGTGTACCGCGCTCTCATGTTGAGGAGGTCAACCGGTTCTCAACGAGATCGCC 780
DB 721 GCGTGTGCGCGCGCTCATGTTGCGAGAGATCAATCGGTTCTCAACGAGATCGCG 780
QY 781 GATGCAACAGTGTACCTGTGCGCACCTCACAGTGGCGCCCGCTGAGGAGCAACAGGTC 840
DB 781 GACGCAACGTTGGGACCGTTCGCGCACTCACGGTGACCCCTGGTGGGCTCAACAGGTT 840
QY 841 CCGGGGAGGTGACATTCACTGACCTGCGTTCCTCGCATGAGAGTGGTTCGCGGTG 900
DB 841 CCGGGGCGGTTCAGTTTCACTGCTCGATCTGCGATCGCCGACGAGGATCGATCCGGCTC 900
QY 901 CTGATCGACGATCTCGGTTCATGTCGCGAGGTGCGCTCCAGCGGTTGGTGGTCC 960
DB 901 CTGGTTCGACAGGATCGAGCGATGTTGGCAGAGTTCGCGCGCGCGCGGAGTCGAGGCC 960
QY 961 GATGTGGATGAATTTTCAATCTCAGCGCGGTGCACTGCTCTACCATGTTGAGCGCC 1020
DB 961 GCGGTGAACGGGTTCTTCGGCTCAGCCCTGTCTGCTCTCGGTGTTGTTGATCGC 1020
QY 1021 GTTTCGGAAGCGGCTTCGCGCTTGCAGTTTCAACACCGGGATATCAGAGTGGGCGGGC 1080
DB 1021 GTGCGGACGCGCGTCCGAACTCGGCTTTCACCATCGCGACATCACAGAGCGGGCAGGG 1080
QY 1081 CAGGACTCGATGTTATCGCCAGGTTCAGGAGTTCGGAATGGTTCGTTTCAAGCCGT 1140
DB 1081 CAGGACTCGATGTTATCGCCAGATCACCGAGTTCGGAATGGTTCGTTTCCCGACGCG 1140
QY 1141 GCTGCGGAGGACCGTTCGCGAAGAAAGGACCGATTTTCGATGCACTTCGCAAGGAACT 1200
DB 1141 GCGGCGGAGGACATGTGCGGAGGAAATGTCGATTTTCAAGATCTCGGAGAGGAGCG 1200
QY 1201 GAGGTTGCTCTCGGGAATGAAGGCACTTGACCGGTA 1239

DB 1201 GATGGTCTTCACGTCGTGACGCGCTTGACCGTGA 1239
RESULT 9
BD176614
LOCUS
DEFINITION
1239 bp DNA linear PAT 18-MAR-2003
BD1 encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid
hydrolase, recombinant DNA, transformed cells, process for
producing protein and process for producing optically active amino
acid.
ACCESSION
BD176614
VERSION
BD176614.1 GI:29122324
KEYWORDS
WO 02072841-A/2.
SOURCE
Microbacterium liquefaciens
ORGANISM
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococcales; Microbacteriaceae; Microbacterium.
REFERENCE
1 (bases 1 to 1239)
Takenaka, Y., Suzuki, S., Onishi, N. and Yokozeki, K.
DNA encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid
hydrolase, recombinant DNA, transformed cells, process for
producing protein and process for producing optically active amino
acid
JOURNAL
Patent: WO 02072841-A 2 19-SEP-2002;
AJINOMOTO CO INC, YASUHIRO TAKENAKA, SHUNICHI SUZUKI, NORIMASA ONISHI,
KENZO YOKOZEKI
COMMENT
OS Microbacterium liquefaciens
PN WO 02072841-A/2
PD 19-SEP-2002
PF 08-MAR-2002 WO 2002JP002173
PR 08-MAR-2001 JP 01P 065814, 27-SEP-2001 JP 01P 298619 PI
YASUHIRO TAKENAKA, SHUNICHI SUZUKI, NORIMASA ONISHI, KENZO PI
YOKOZEKI
PC C12N15/55, C12N1/15, C12N1/19, C12N1/21, C12N5/10, C12N9/14, C12N9/
PC 86, C12P13/04,
PC C12P13/22, C12P41/00, C12N15/52
CC Chase
FH Key
FT CDS
Location/Qualifiers
(1). (1239).
source
1..1239
/organism="Microbacterium liquefaciens"
/mol_type="genomic DNA"
/db_xref="taxon:33918"
ORIGIN
Query Match 59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity 74.8%; Pred. No. 4e-120;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTGCAGAAAGCGCAGCGCGCATTTGAGAAAGAGATCGGGAGCTCTCCCCG 60
DB 1 GTGACCTGCACAGCGCGCGCGCATCGCATCGAGGAGGAGCTTGGACTCTCTCCCCG 60
QY 61 TTCTCGGAGAAAGCGCGCGCTTTACCGCGCTGACCTACATCCAGAGATGCGCGCGG 120
DB 61 TTCTCGGTCGAAGGCGCGCGGTGACAGCTCTCACGTACATCCGAGACGCGCGCGG 120
QY 121 CGGGAACCGCTCATTTGCGGCTATGAAAGCGCGCGCTTGAGCGTTCGTGAAGACGCATC 180
DB 121 CGAGAGGTGATCGTCCGCGCATGCGAGCGGACGGGCTGAGCGTCCACGAGGACGCTCTC 180
QY 181 GGAAACATCATCGGCGGAGTGAAGCACTGATCCGAGCTTCCTCGGATCCGGTCCGT 240
DB 181 GGCAACATCATCGTCCGGCTGAGGGAGCGAGCCCGCTCTCGCGCGATCGCTTCGGC 240
QY 241 TCACACTTCGATTCCTCCGAAACCGCGGATGTTTGTGAGCACTGACGCGGTGTGTGC 300
DB 241 TCGCACTTCGACTCGTTCGCAACCGCGGATGTTTCGACGGCACCGCGGCGTGTGTGC 300
QY 301 GCGCTTAGCGCTCCCGGCTGATGTCGAGAGCGGCTACGTCAGTGAATCGGATCCATTTAG 360

Db 301 GCGCTCGAGGCTCGCAGGCTGCTCGAGGAGCGGATGTGAACCGTCATCCTCTCGAG 360
Qy 361 TTCATCGCGATCGTGAGAGAAAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCG 420
Db 361 GTCATCGCGATCGTCGAGAGAGGGCAACCCGCTTCAGCAGCGGCATGCTGGCGGCTCGC 420
Qy 421 GCCATTGCGAGGCTGGTCCGCACAGGAACTGGACTCTTTGGTTGATGAGGATGGAGTG 480
Db 421 GCGATCGCGGGCTGCTGTCGACGCCGATCTGGACACCTGGTGGACGAGCGGCGTG 480
Qy 481 TCCGTTAGCGAGCGGCTACTGCTTCGCTTTGAAGCCGGCGAACTGCAAGCTCGACGC 540
Db 481 ACGGTGCGAGCGCGCCACAGCCCTTCGCGCTTGAACCGGTGAGCTGCGAGCGCGGCC 540
Qy 541 CGCTCGCGCGGACCTGGGTGCTTTTATCGAACTACACATTTGAACAAGACCGATCTC 600
Db 541 CGTACGAGGGATGACCTTCGCGCCCTTCATCGAGTTGCAATCGACGAGCGGCGCATCTC 600
Qy 601 GAGCAGGAGCAATAGAGATCGAGTTGTAACCTCCATGTTGGCGTTTCGCGCATTTGGG 660
Db 601 GAGCAGGAGAGAGGTGAGATCGCGCTGTGACGGGGATCGTCGTTGCGCGCTTCGG 660
Qy 661 GTTGCCGTCAAAGGAGAGGACACACGCGGCACAAACCCCATGCACTGCGCCAGGAT 720
Db 661 ATCAGGTGGAGGCGAGGAGCACACGCGGGACGACCCCATGCACTGCGCGCAGGAC 720
Qy 721 GCGCTGGTACCGCGCTCTCATGTGAGGAGGTTCAACCGGTTGTTGTTCAACGAGATGCC 780
Db 721 GCGCTGGTGC CGCGCGGCTCATGTTGCGAGAGATCAATCGGTTGTTCAACGAGATCGG 780
Qy 781 GATGCGACAGTGGCTTACGTTGGCCACCTCACAGTGGCCCGCGGTGGAGGCAACCAAGTTC 840
Db 781 GAGCGACAGTGGCGACCTGCGCCACCTCACGTTGACCCCTGGTGGCTCAACCAAGTTC 840
Qy 841 CGCGGGGAGGTGACTTCACACTGGAACCTGCGTTCTCGCGATGAGGAGTTCGTCGCGTG 900
Db 841 CCGCGGGGCTCGAGTTCAAGCTCGATCTGCGATCGCCACAGGAGTTCGATCGCGTTC 900
Qy 901 CTGATCGACCGCATCTCGGTATGTTGCGGAGGTCGCTCCAGCGCGGTGTCGCTGCC 960
Db 901 CTGGTCGACAGGATCGAGCGATGTTGGCAGAACTGCGCGCGCGCGGCGGAGTTCGAGGCC 960
Qy 961 GATGTGGATGAATTTTCAATCTCAGCCGCGTGCAGCTGGCTTCCTACCATGTTGACGCC 1020
Db 961 GCGGTGAACGGGTTCTTCGCGCTCAGCCCTGTCGCTCTGCTTCGCGTGGTCTGGATCGC 1020
Qy 1021 GTTCGCGAAGCGGCTCGGCTTCAGTTTCACACCCGGATATCAGCAGTGGGCGCGGC 1080
Db 1021 GTGCGGAGCGGCTCGGAACTCGGCTTCACCCATCGCGCATCACAGCGGGGCGAGG 1080
Qy 1081 CAGGACTCGATGTTTCATCGCCAGGTACGAGCGTTCGGAATGGTTTTCGTTCCAGCGGT 1140
Db 1081 CAGGACTCGATGTTTCATCGCCAGATCACCGACGTTCGGAATGGTTTTCGTTCCAGCGGC 1140
Qy 1141 GCTGCGCGGAGCACGTTTCCGAAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACT 1200
Db 1141 GCCGGCGGAGCCATGTGCGGAGGAATGCTCCGATTTTCGAGATCTCGCGNAGGGAGG 1200
Qy 1201 GAGGTTGCTCTCCGGGTAAATGAAGCACCTTGACCGGTAA 1239
Db 1201 GATGTGGTCTTCACGTCGTGACGCGGCTTGACCGGTGA 1239

RESULT 10
BD180889 1239 bp DNA linear PAT 15-MAY-2003
LOCUS BD180889
DEFINITION DNA encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid hydrolase, recombinant DNA, transformed cells, process for producing protein and process for producing optically active amino acid.
ACCESSION BD180889
VERSION BD180889.1 GI:30791807
KEYWORDS JP 2002330785-A/2.

SOURCE Microbacterium liquefaciens
ORGANISM Microbacterium liquefaciens
REFERENCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
AUTHORS 1 (bases 1 to 1239)
TITLE Takenaka,Y., Suzuki,S., Onishi,N. and Yokozeki,K.
DNA encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid hydrolase, recombinant DNA, transformed cells, process for producing protein and process for producing optically active amino acid
JOURNAL Patent: JP 2002330785-A 2 19-NOV-2002;
COMMENT AJINOMOTO CO INC
OS Microbacterium liquefaciens
PN JP 2002330785-A/2
PD 19-NOV-2002 JP 2001298619
PF 27-SEP-2001 JP 2001298619
PI YASUHIRO TAKENAKA,SHUNICHI SUZUKI,NORIMASA ONISHI,KENZO PI YOKOZEKI
PC C12N15/09,C12N1/15,C12N1/19,C12N1/21,C12N5/10,C12N9/78,C12P13/22,
PC C12P41/00/(C12N1/21,C12R1:01),(C12N9/78,C12R1:01),(C12P41/00,
PC C12R1:01),
PC C12N15/00,C12N5/00
CC Chase
FH Key Location/Qualifiers
FT CDS (1)..(1239).

FEATURES
source
1..1239
/organism="Microbacterium liquefaciens"
/mol_type="genomic DNA"
/db_xref="taxon:33918"

ORIGIN
Query Match 59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity 74.8%; Pred. No. 4e-120;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;

Qy 1 GTGACCTTCGAGAAAGCGCAGCGCGCATTCGAGAAAGAGATCCGGGAGCTCTCCCGG 60
Db 1 GTGACCTTCGAGAGCGCGCGCGCATTCGAGAAAGAGATCCGGGAGCTCTCCCGG 60
Qy 61 TTCTCGGAGAGCGCGCGGTGTTTACCGCGGTGACCTACCTCCAGAGATGCGCGCGCG 120
Db 61 TTCTCGGTCGAGAGCGCGCGGTGACAGCTCTACGTACACTCCGGAGACGCGCGCGG 120
Qy 121 CGGAAACGCTCATTCGGGCTATGAAGCGCGCGCTTGAGCGTTCGTGAAGACGACATC 180
Db 121 CGAGAGGTGATCGTCGCGCCATGACGCGAGCGGGGCTGAGCGTCCACGAGAGCGCTCTC 180
Qy 181 GGAAACATCATCGCGCGGAGCGTGAAGGCACTCATCGGAGCTTCCTGCGATCGCGTCCGT 240
Db 181 GGCAACATCATCGTGGCGGTGAGGGAGCGACCCCGCTCTGCGGCGATCGCCTTCGCG 240
Qy 241 TCACACTTCGATTCTGTCGAAACCGCGGGATGTTTGGATGGCACTGCAAGGCGTGGTGTGC 300
Db 241 TCGCACTTCGACTCGTTCGCAACCGCGGGATGTTTCGACGCGCACCGCGGGCGTGGTGTGC 300
Qy 301 GCCCTTGAGGCTGCCGGGTGATGCTCGAGAAACGGCTACGTTGAATCGGCATCCATTTCAG 360
Db 301 GCGCTCGAGGCTGCGAGGGTCTCAGAGAGCGGATATGTGAACCGCTCATCTCTCTCGAG 360
Qy 361 TTCATCGGATCGTGGAGGAGGAGGCGCCGCTTCAGCAGTGGCATGTTGGCGCGCGCG 420
Db 361 GTCATCGGATCGTCGAGAGAGGAGGCGCACCGCTTCAGCAGCGGATGCTGGCGGTGCG 420
Qy 421 GCCATTGAGGTTGGTTCGCGCAGGAACTGGAATCTTTGGTTGATGAGATGAGATG 480
Db 421 GCGATCGCGGGCTCGTGTCCGACCGCCATCTGGACACCCCTGGTGGACGAAAGACGCGGTG 480
Qy 481 TCCGTTAGGAGGCGGCTACTGCTTCGGCTTGAGCGGCGGAGCTGACGAGTGCAGCC 540
Db 481 ACGGTGCGGAGCGGCGCACCGCCCTTCGGGCTTGAAACCGGGTGAGCTCGCGAGCGCGCC 540

QY 541 CGCTCCGGCGGACCTCGTGTCTTTTATCGAACTACATGTAACAAGACCGATCCTC 600
Db 541 CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCACATCGAGCGGGCGATCCTC 600
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCGTTGGCGTTGCGCATTCGG 660
Db 601 GAGCAGGAGAGTGGAGATCGCGTCTGTCACGGGGATCGTGGTGTCCGCGCTTCGG 660
QY 661 GTTGGCGTCAAGGCGAGAGCGACCAACCGCGGACCAACCCCATGACCTCGGCCAGGAT 720
Db 661 ATCAGCGTGGAGGCGAGGAGCGACCAACCGCGGAGCACCCCATGACCTCGGCCAGGAC 720
QY 721 GCGCTGGTACCCCGCTCTCATGTTGAGGAGTCAACCGGTTCTCAACGAGATCGCC 780
Db 721 GCGCTGGTGGCGCGCGCTCATGGTGGAGAGATCAATCGGTTCTGTCACAGGATCGCG 780
QY 781 GATGGCAGAGTGGCTTACCGTTGGCCACCTCACAGTGGCGCCCGGTTGGAGCAACAGGTC 840
Db 781 GACGGACCGTGGCGACCGTGGCGACCTCACGGTGAACCCCTGGTGGGCTCAACAGGTT 840
QY 841 CCGGGGAGGTGGAATTCACTGAGACCTCGCTTCCGATGAGGAGTGGCTTCCGCGTG 900
Db 841 CCGGGGGCGTGGATTCAGCTCGATCTGGATCGCCACGAGGAGTCGATCCGCGTC 900
QY 901 CTGATCGACCGCATCTCGGTATGTTGGCGGAGGTGCGCTCCAGCGCGGTGGTGGCTGCC 960
Db 901 CTGGTGCACAGGATCGAGGCGATGGTGGCAGAGTGGCGCGCGCGCGGAGTCGAGGCG 960
QY 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGACGTGGCTCCTACCATGGTGGAGCGC 1020
Db 961 GCGTGAACGGGTTCTTCGGGCTCAGCCCTGTCGGTCTGTCCTCGGTGGTGGATCGC 1020
QY 1021 GTTCCGGAAGCGGCTTCGGCTTTCAGATTCAACACCGGATATCAGCAGTGGGGCGGGC 1080
Db 1021 GTGCGGACGCGGCGTCCGAATCTCGGCTTCAACCATCGGACATCAGAGCGGGCAGG 1080
QY 1081 CAGCACTCATGTTATCGCCAGGTACGAGAGTCGGAATGGTTTTGTTTCCAAAGCGT 1140
Db 1081 CAGCACTCATGTTATCGCCAGATCAGCAGTCGGAATGGTTTCCGTCGCCAGCGCG 1140
QY 1141 GCTGGCGGAGCACCTTCCGAGAAATGACCGATTTCGATGACCTTCGCAAGGAAT 1200
Db 1141 GCGGGCGAAGCATGTGCGGAGGAATGGTCCGATTCGACGATTCGCGAAGGGAGCG 1200
QY 1201 GAGGTTGTCTTCGGGTAATGAGGCACTTGACCGGTAA 1239
Db 1201 GATGTTCTTTCAGCTCGTACGGCGCTTGACCGGTGA 1239

RESULT 11
LOCUS BD181025 1239 bp DNA linear PAT 15-MAY-2003
DEFINITION 5-Substituted hydantoin racemase, DNA encoding the same, recombinant DNA, transformed cells and process for producing optically active amino acid.
ACCESSION BD181025
VERSION BD181025.1 GI:30791943
KEYWORDS JP 2002330784-A/3.
SOURCE Microbacterium liquefaciens
ORGANISM Microbacterium liquefaciens
REFERENCE Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.
AUTHORS Suzuki, S., Onishi, N. and Yokozeki, K.
TITLE 5-Substituted hydantoin racemase, DNA encoding the same, recombinant DNA, transformed cells and process for producing optically active
JOURNAL Patent: JP 2002330784-A 3 19-NOV-2002; AJINOMOTO CO INC
COMMENT OS Microbacterium liquefaciens
PN JP 2002330784-A/3
PD 19-NOV-2002
PF 13-SEP-2001 JP 2001278739

PI SHUNICHI SUZUKI, NORIMASA ONISHI, KENZO YOKOZEKI PC
C12N15/09, C12N1/21, C12N9/90, C12P13/02, C12P13/04, C12P17/10, PC
C12N15/00
CC Chase
FH Key
FT CDS
Location/Qualifiers
(1). (1236).
source
1. 1239
/organism="Microbacterium liquefaciens"
/mol_type="genomic DNA"
/db_xref="taxon:33918"
FEATURES
ORIGIN
Query Match 59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity 74.8%; Pred. No. 4e-120;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCG 60
Db 1 GTGACCTTCAGAGCGCGCGCGCATTCGATCGATCGAGGAGGCTCTGGACTCTCTCCCG 60
QY 61 TTCTCGGAGAAAGCGCGCGCTGTACCGGCTGACCTACATCCAGAGCATGCCGCGCG 120
Db 61 TTCTCGGTCAGAGGCGCGCGCTGACAGCTCTACATACATCCGGAGACGCGCGCGCG 120
QY 121 CCGGAAACGCTCATTTCCGGCTATGAAAGCGCGCGCTTGAGCGTTCTGGAAGACGACTC 180
Db 121 CGAGAGGTGATCGTCGCGCCATGACGCGACGGGCTGAGCGTCCACGAGGAGCTCTC 180
QY 181 GAAACATCATCGCGGAGTGAAGCACTGATCCGAGCTTCTTCGATCGCGGTCCGT 240
Db 181 GCGCAACATCATCGTCGCGGTGAGGGAGCGACCCGCTCTCCGCGCATCGCCTTCGC 240
QY 241 TCACACTTCGATCTGTTCGAAACGCGCGGATGTTTGTGACACTGACGCGGTGGTGC 300
Db 241 TCGACTTCGATCGTTCGCGAAACGCGCGGATGTTTCGACGCGACCGCGCGGTGGTGC 300
QY 301 GCGCTTGGGCTCGCGGCTGATGCTGGAGAACGCGTACGTAATCGGATCCATTGAG 360
Db 301 GCGCTCGAGGCTCGAGGCTGCTCGAGGAGAGCGGATATGTGAACGCTCATCTCTCGAG 360
QY 361 TTCTATCGCATCGTGGAGGAGAAAGGCGCGCTTCAGAGTGGCATGTTGGCGCGCG 420
Db 361 GTCATCGCATCGTGAAGAGAGGAGCGACCCGCTTCAGCAGCGGCTGCTGGCGGTGC 420
QY 421 GCGATTGACAGGTTGCTCGCCGACAGGAACTGGACTCTTTGTTGATGAGATGAGAGTG 480
Db 421 GCGATCGCGGGCTCGTGTGACGCGGATCTGGACACCTCGTGGACGAGACGCGGTG 480
QY 481 TCCGTTAGCGAGCGGCTACTGCTTCGCTTTGAAAGCGCGGCGAACTGCGAGGTGCGAGCC 540
Db 481 ACGGTGCGGAGCGCGCCACGCGCTTCGCGCTGGAACCGGGTGAGCTCGGAGCGCGCC 540
QY 541 CGCTCGGCGCGGACTGCGTGTCTTTATCGAACTTACATTTGAACAAGAGACGATCTCTC 600
Db 541 CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCAATCGAGCAGGCGCGGATCTCTC 600
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACCTCCATCGTTGGCGTTGCGCATTCGG 660
Db 601 GAGCAGGAGAGGTGGAGATCGCGTCTGTCGAGGAGATCGTGGTGTCCGCGCTTCCCG 660
QY 661 GTTGGCGTCAAGGCGAGAGCGACCAACCGCGGACCAACCCCATGACCTCGGCCAGGAT 720
Db 661 ATCAGCGTGGAGGCGAGGAGCGACCAACCGCGGAGCGACCCCATGACCTCGGCCAGGAC 720
QY 721 GCGCTGGTACCGCGCTTCTCATGTTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCC 780
Db 721 GCGCTGGTGGCGCGCGCTCATGGTGGAGAGATCAATCGGTTCTGTCACAGAGATCGCG 780
QY 781 GATGGCAGAGTGGCTTACCGTTGGCCACCTCACAGTGGCGCGCGGTGGAGGACGAGGTC 840
Db 781 GACGGCAGGTTGGCGACCGTTCGCGCGCACCTCACGGTGGTGGTGGTTCACAGGTT 840

```
QY      841  CCGGGAGGTGAGCTTACACTGACCTGCGCTTCTCGCATGAGAGTGCCTCCGCTG  900
Db      841  CCGGGGGCGTGCAGTTTACGCTCGATCTGCCATCGCCACGAGAGTGCATCCGGCTC  900

QY      901  CTGATCGACCGCATCTCGGTCATAGTTCGGCGAGGTGCGCTCCAGCCGGTGTGGCTGCC  960
Db      901  CTGGTCGACAGGATCGAGGCGATGTTGGCGAGAAGTCGCCGCCGCGGCGGAGTCGAGGCC  960

QY      961  GATGTGATGAATTTTCAATCTCAGCCCGGTGCGAGCTGGCTCCCTACCATGTTGAGCGCC  1020
Db      961  GCGGTGAACGGGTTCTTCGCGCTCAGCCCTGTCGGTCTGTCTCTCCGGTGTGCTGGATCGC  1020

QY      1021  GTTCCGGAAGCGGCTTCGSCCTTCGAGTTTACACACCGGATATCAGCAGTGGGCGGCGC  1080
Db      1021  GTGCGGACCGCGGTCCGAATCTCGCTTCGGCTTACCCTACCGCATCAGAGCGGGCAGGG  1080

QY      1081  CACGACTCGATTTTCAATCGCCAGGTTCAGCGCTCGGAATGGTTTTCGTTCCAAAGCCGT  1140
Db      1081  CACGACTCGATGTTTCAATCGCCAGATCACCGAGTCCGAATGGTGTTCGTTCCCAAGCCGC  1140

QY      1141  GCTGCGCGGAGCGATGTTCCCGAAGATGGACCGATTTGATGACCTTCGCAAGAACT  1200
Db      1141  GCGCGCGGAAGCATGTGCGGAGGAATGTTCCGATTTTCAGCATCTGCGGAAGGGGACG  1200

QY      1201  GAGGTTGCTCCCGGTAATGAAGCAGCTTGACCGGTAA  1239
Db      1201  GATGTGGTCTTACGCTGTGACGGCTTTGACCGGTGA  1239
```

RESULT 12

```
AX397883
LOCUS      AX397883          3343 bp      DNA      linear      PAT 27-MAY-2002
DEFINITION Sequence 7 from Patent EP1188826.
ACCESSION AX397883
VERSION   AX397883.1  GI:21260750
```

KEYWORDS

Microbacterium liquefaciens
Microbacterium liquefaciens
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Micrococccineae; Microbacteriaceae; Microbacterium.

REFERENCE

1 Suzuki, S.C., Onishi, N.C. and Yokozeki, K.C.
5-substituted hydantoin racemase, dna coding for the same, and
process for producing optically active amino acids
Patent: EP 118826-A 7 20-MAR-2002;
Ajinomoto Co., Inc. (JP)

FEATURES

Location/Qualifiers
1..3343
/organism="Microbacterium liquefaciens"
/mol_type="unassigned DNA"
/db_xref="taxon:33918"

ORIGIN

```
Query Match      59.7%; Score 739.8; DB 6; Length 3343;
Best Local Similarity 74.8%; Pred. No. 3.5e-120;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;

QY      1  GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAAAGAGATCCGGAGCTCTCCGG  60
Db      2105  GTGACGCTGACAGCGCGCGGCCGATCGATCGAGGAGGAGCTCTGGACTCTCTCCGCG  2164

QY      61  TTCTCGGAGAGGCCCGGTGTACCGGCTGACCTACCTCCAGAGCATGCCCGCGG  120
Db      2165  TTCTCGGTGAAAGGCGCGCGGTGACAGCTCTACGTACACTCCGAGACGCGCCGCG  2224

QY      121  CCGGAAACGCTCATTTGCGGCTATGAAAGCGCGCTTGAGCGTTTCGTGAAGACGACTC  180
Db      2225  CGAGAGGTGATCGTCCGCCCATGACGCGGACGGGGCTGAGGCTCCACGAGGACGCTCTC  2284

QY      181  GGAACATCATCGGCGGACGTGAAGGCACTGTATCCGGAGCTTCTTGGCATCCGGTCCGT  240
Db      2285  GGCAACATCATCGGTCCGGCTGAGGGAGCGACCCCGCTCTGCGCGCATCGCTTCCGCG  2344
```

RESULT 13

```
BD176615
LOCUS      BD176615          3343 bp      DNA      linear      PAT 18-MAR-2003
DEFINITION DNA encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid
```

```
QY      241  TCACATCTCGATCTGTCCGAAACGGCGGATGTTTGATGGCACTGACAGGCGTGGTGGC  300
Db      2345  TCGCACTTCGACTCGGTCCGCAACGGCGGATGTTTCGACGGCAACGGCGGCGTGGTGGC  2404

QY      301  GCCCTTAGGCTGCCCGGTGATCTGGAGAAACGGCTACGTAATCGGCATCCATTTGAG  360
Db      2405  CGCTCGAGGCTCGAAGGGTGTCTCAGGAGAGCGATATGTGAACCGCTCATCTCTCGAG  2464

QY      361  TTCAATCGGCTGCTGGAGAGGAGGGCCCGCTTCAGACGTGGCATGTTGGGGCGCGCG  420
Db      2465  GTCAATCGGATCTGTCGAAGAGGAGGCGACCCGCTTCAGACGCGGATGCTGGGCGGTGCG  2524

QY      421  GCCATTGACAGGTTGGTTCGCCGACAGGAACTGGACTCTTTGGTTGATGAGGATGGAGTG  480
Db      2525  GCGATCGCGGGCTCTGTTCGACGCCGATCTGGACACCTTGGTGGACGAGACGGCGGTG  2584

QY      481  TCCGTTAGGCAAGCGGCTTATCTGCTTCGCTTGAAGCCGGCGGCAACTGCAAGGCTGACGC  540
Db      2585  ACGGTGCGGAGGCGGCGCACCGGCTTCGGGCTGGAAACGGGCTGAGCTGCGACGCGCGCC  2644

QY      541  CGCTCCGCGCGGACCTCGGTGCTTTTATCGAACTACACATTGAACAGGACCGATCCCTC  600
Db      2645  CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCACATCGAGACAGGGGCGCATCTCTC  2704

QY      601  GAGCAGGAGCAATAGAGATCGGAGTTGTAACTCCATCTGTTGGCGTTTCGCGCATTTGCGG  660
Db      2705  GAGCAGGAGAGGTGGAGATCGGCTGTGACGGGGATCGTCSGTGTCGCGCTTCCGG  2764

QY      661  GTTCCCGTCAAAGCGAAGCGACCAACCGCGCAAAACCCCATGCGACTCGGCCAGGAT  720
Db      2765  ATCACGGTGGAGGCGAGGAGCGACCAACCGCGGACGACCCCATGCGACTCGCGGAGGAC  2824

QY      721  CGCTGTGTACCCGCGCTCTCATGTTGAGGAGGCTCAACCGGTTTCGTAACAGAGATCGCC  780
Db      2825  CGCTGTGTGCGCGCGCGCTCATGTTGAGAGATCAATCGGTTCTGTCAAAGAGATTCGCG  2884

QY      781  GATGCGACAGTGGCTACCGTTGGCCACCTCACAGTGGCGCCCGCGGTGGAGGCAACAGAGTC  840
Db      2885  GACGCGACGTTGGGACCGTGGCCACCTCACGTTGACCCCTGTGTGGGCTCAACAGGTT  2944

QY      841  CCGGGGAGTGGACTTACACTCGACTCGCTGCTTCTCGCATGAGGAGTGCCTCCCGTG  900
Db      2945  CCGGGGCGTTCGAGTTTACGCTCGATCTGCGATCGCCCGCACGAGGAGTTCGATCCGGCTC  3004

QY      901  CTGATCGACCCGATCTCGGTGATGTCGCGCAGGTGCGCTCCAGCGCGGTGGCTGCC  960
Db      3005  CTGTTCGACAGGATCGAGGCGATGGTGGCAAGTTCGCCCGCGCGCGGAGTTCGAGGCC  3064

QY      961  GATGTGATGAATTTTCAATCTCAGCCCGGTGAGCTGGCTCTTACCATGTTGACGCGCC  1020
Db      3065  GCGGTGAACGGGTTCTTCGCGCTCAGCCCTGTCCGTTCTGTCTCCGGTGTGTTGATCGC  3124

QY      1021  GTTCGGAAGCGGCTCGGCCCTTTCAGTTTCAACACCGGGATATCAGCAGTGGGGCGGGC  1080
Db      3125  GTGCGGACGCGCGGCTCCGAACTCGGCTTCAACCATCGCGACATCACAGCGGGGCGAGGG  3184

QY      1081  CACGACTCGATGTTTCATCGCCCGGAGGTTCAGGACGTCGGAATGGTTTTCGTTCCAGCCGT  1140
Db      3185  CACGACTCGATGTTTCATCGCCCGGAGTACCAGACGTGCGGAATGGTTTTCGTTCCCGGCGC  3244

QY      1141  GCTGCCCGGAGCCAGTTCGCCGAAAGATGGACCGATTTTCGATGACCTTCGCAAGGAACT  1200
Db      3245  GCCGGGCGAAGCCATGTCCCGGAGGATGGTCCGATTTTCGACGATCTCGCGGAGGGGAGC  3304

QY      1201  GAGGTTGCTCTCCGGGTAATGAAGGCACTTGACCGGTAA  1239
Db      3305  GATGTGTCTTCTTACGCTCGTACGCGGCTTGACCGGTGA  3343
```


REFERENCE	Micrococciaceae; Microbacteriaceae; Microbacterium.	Db	2585	ACGGTGCAGGCGGCGCCACCGCTTCGGGCTGGAAACCGGCTGAGCTCGCGACGCGCGCC	2644
AUTHORS	1 (bases 1 to 3343)				
TITLE	Takenaka,Y., Suzuki,S., Onishi,N. and Yokozeki,K. DNA encoding hydantoinase, DNA encoding N-carbamyl-L-amino acid hydrolase, recombinant DNA, transformed cells, process for producing protein and process for producing optically active amino acid	Qy	541	CGCTCCGCGCGGACCTCGCTGCTTTATCAAACTACACATTGAACAGGACCGATCCTC	600
JOURNAL	Patent: JP 2002330785-A 3 19-NOV-2002;	Db	2645	CGTACAGGGATGACCTTCGCGCTTCATCGATTGACATCGAGCGGGGCCGATCCTC	2704
COMMENT	AJINOMOTO CO INC OS Microbacterium liquefaciens PN JP 2002330785-A/3 PD 19-NOV-2002 PF 27-SEP-2001 JP 2001298619 PI YASUHIRO TAKENAKA, SHUNICHI SUZUKI, NORIMASA ONISHI, KENZO PI YOKOZEKI	Qy	601	GAGCAGGAGCAATAGAGATCGGAGTTGTAACTCCATCTGTTGGCTTCGGCATTTGCGG	660
		Db	2705	GAGCAGGAGAGGTGGAGATCGGCTCGTACGGGGATCGTCGGTGTCCGCGCTTCGG	2764
		Qy	661	GTTGCCGCTCAAGCAGAGCAGCAGCGCGGCAAAACCCCATGACACTCGCGCAGAT	720
		Db	2765	ATCACGCTGGAGGCGCAGGAGCGACAGCCCGGAGCAGCCCATGACACTCGCGCAGGAC	2824
		Qy	721	CGCTGTGTACCCCGCTCTCATGTTGAGGAGGTCAACCGGTTTCGTCACAGAGATCGCC	780
		Db	2825	CGCTGTGTGCGCGCGGCGCTCATGTTGCGAGAGATCAATCGGTTTCGTCACAGAGATCGC	2884
		Qy	781	GATGGCACAGTGGCTACCGTTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACACAGGTC	840
		Db	2885	GACGGCAGGTGGGACCGCTCGGCCACCTCACGTTGACCCCTGGTGGGCTCAACAGGTT	2944
		Qy	841	CCGCGGAGGTGGACTTCACACTGGAACCTGCGTTCTTCGCGATGAGGAGTCTCGCGGTG	900
		Db	2945	CCCGGGGCGGTCGAGTTTCACTGCTGATCTGCGATCGGCCACGAGGAGTCTCGCGCTC	3004
		Qy	901	CTGATCGACCGCATCTCGGTCATGTCGGCGAGGTGCGCTCCAGGCGGTTGGCTGCC	960
		Db	3005	CTGTCGACAGGATCGAGGCGATGGTGGCAAACTCGCGCGCGCGGCGGAGTCGAGGCC	3064
		Qy	961	GATGTGATGAATTTTCAATCTCAGCCCGGTGCGAGTGGCTCTTACCATGGTGGACGCC	1020
		Db	3065	CGGTGAAACGGTTCTTCGCGCTCAGCCCTGTCGGTCTGTCTCCGGTGGTCTGGATCGC	3124
		Qy	1021	GTTCCGGAAGCGGCTCGGCTTCGAGTTTACACACCGGGATATCAGCAGTGGGCGGCG	1080
		Db	3125	GTGCGCAGCGGCGGTCGAACTCGGTTTACCCATCCGACATCAGAGCGGGGCGAGGG	3184
		Qy	1081	CACGACTCGATGTTTCGCGCGGCGGCTCAGGACCTCGGATGTTTCTGTTCCAGCCGT	1140
		Db	3185	CACGACTCGATGTTTCGCGCGGCGGCTCAGGACCTCGGATGTTTCTGTTCCAGCCGT	3244
		Qy	1141	GCTCGCGGAGCCACGTTTCCGAAAGATGGAACCATTTTCGATGACCTTCGCAAGGAACT	1200
		Db	3245	GCGCGGAGACCATGTGCGGAGGATGTTCCGATTCGACGATCTCGCGAGGAGGACG	3304
		Qy	1201	GAGTTGCTTCGCGGTAATGAAGCAGCTTGAACCGGTAA	1239
		Db	3305	GATGTGCTCTTACGTCGTGACGCGCTTGACCGGTGA	3343
RESULT 15					
BD181026					
LOCUS	BD181026				
DEFINITION	5-Substituted hydantoin racemase, DNA encoding the same, recombinant DNA, transformed cells and process for producing optically active amino acid.				
ACCESSION	BD181026				
VERSION	BD181026.1				
KEYWORDS	JP 2002330784-A/4.				
SOURCE	Microbacterium liquefaciens				
ORGANISM	Microbacterium liquefaciens				
REFERENCE	Bacteria: Actinobacteria; Actinobacteridae; Actinomycetales; Micrococciaceae; Microbacteriaceae; Microbacterium.				
AUTHORS	1 (bases 1 to 3343)				
TITLE	Suzuki,S., Onishi,N. and Yokozeki,K. 5-Substituted hydantoin racemase, DNA encoding the same, recombinant DNA, transformed cells and process for producing optically active				
JOURNAL	Patent: JP 2002330784-A 4 19-NOV-2002;				
COMMENT	AJINOMOTO CO INC OS Microbacterium liquefaciens PN JP 2002330784-A/4				

```
PD 19-NOV-2002
PF 13-SEP-2001 JP 2001278739
PI SHUNICHI SUZUKI,NORIMASA ONISHI,KENZO YOKOZEKI PC
C12N15/09,C12N1/21,C12N9/90,C12P13/02,C12P13/04,C12P17/10, PC
C12N15/00
CC 5-Substituted hydantoin racemase, DNA encoding the same, CC
recombinant DNA,
CC transformed cells and process for producing optically active
CC amino acid
FH Key Location/Qualifiers
FT source 1. 3343
FT /organism='Microbacterium liquefaciens'.
FEATURES
source
1. 3343
/organism='Microbacterium liquefaciens'
/mol_type='genomic DNA'
/db_xref='taxon:33918'
ORIGIN
Query Match 59.7%; Score 739.8; DB 6; Length 3343;
Best Local Similarity 74.8%; Pred. No. 3.5e-120;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTCGAAGCGCAAGCGCGCGCATTCAGAAAGAGATCCGGGAGCTCTCCGG 60
Db 2105 GTACGCTGAGCAGCGCGGCGCGATCGCATCGAGGAGAGCTCTGGAATCTCTCCC 2164
QY 61 TTCTCGGCAAGAGCGCCCGGTGTACCGGCTGACCTACACTCCAGAGATGCGCGCG 120
Db 2165 TTCTCGGTCGAAGGCGCGCGGTGACAGCTCTACGTCACGTCACGTCGAGGAC 2224
QY 121 CGGGAACGCTCATTCGGGTATGAAAGCGCGCGCTTGAGCGTTTCGTGAAGACGAC 180
Db 2225 CGAGAGGTGATCTCGCGCGCATGACGCGAGCGGCGTGAGCGTCCACGAGGACG 2284
QY 181 GGAAACATCATCGGCGAGCGTGAAGCACTGATCCGGAGCTTCTCGGATCGCGGT 240
Db 2285 GGCAACATCATCGGTGCGCGGTGAGGGAGCGACCCCGCTCTGCGCGGATCGCCT 2344
QY 241 TCACATTCGATTCGTCCGAAACCGCGGATGTTGTGACACTCGAGCGGTGGTGTG 300
Db 2345 TCGCATTCGATTCGTCGCAACCGCGGAGTGTTCGACGGCACCGCGGCGGTGTG 2404
QY 301 GCCCTTGAGGCTGCCCGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTG 360
Db 2405 CGGCTCGAGCTCGGAGGTGCTGACAGGAGCGGATATGTGAACCGTCACTCTCG 2464
QY 361 TTCATCGCATCTGAGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCG 420
Db 2465 GTCATCGCATCTGCAAGAGGAGGAGCACCCGCTTCAGCAGCGGCGATGCTGGG 2524
QY 421 GCCATTGACAGGTGGTTCGCGGACAGGAACTGGACTCTTTGGTTGATGAGGATG 480
Db 2525 GCGATCGCGGGTCTGTGTCCGACGCGGATCTGGACACCTGTGTGACGAGCGGT 2584
QY 481 TCCGTTAGCAGCGGCTACTGCTTCGCTTGAAGCCGCGGCGAACTGACAGGCTGAG 540
Db 2585 ACGGTGCGAGCGCGCACGCGCTTCGCGCTGGAACCGGCTGAGCTGCGGACGG 2644
QY 541 CGCTCCGCGCGGACCTCGGTGCTTTTATCGAACTACACATTTGAACAGGACCGAT 600
Db 2645 CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCACATCGAGCAGGCGCGAT 2704
QY 601 GAGCAGGAGCAATAGAGATCGAGTGTGAACCTCATCTGCGGTTCGCGCATTCGG 660
Db 2705 GAGCAGGAGAGGTGAGATCGCGCTGTCGACGGGATCGTGTGTCGCGCTTCGG 2764
QY 661 GTTGCCGTCGAAGCGAGAGCGACCGCGGACAAACCCCATGCACTGCGCCAGGAT 720
Db 2765 ATCAGCGTGGAGGCGAGGAGCGACCGCGGACGACCCCATGCACTGCGCGAGG 2824
QY 721 CGGCTGGTACCCGCGCTCTCATGTGAGGAGGTCAACCGGTTCTGCAACGAGATCG 780
|||||
```

```
Db 2825 GCGTGGTCCCGCGCGCTCATGTGTCGAGAGATCAATCGGTTCTGTCACAGAGATCGG 2884
QY 781 GATGGCACAGTGGCTACCGTTGGCCACCTTCACAGTGGCCCCCGGTGGAGGCAACCAAGGTC 840
Db 2885 GACGCGACGCTGGCGACCGTTCGCGCACCTTCACGCTGACCCCTGGTGGGCTCAACCAAGGTT 2944
QY 841 CCGGGGAGGTGAGCTTCACACTGACCTGCGTTCCTCCGCATGAGGAGTGGTCCCGCGT 900
Db 2945 CCGGGGCGTGCAGTTACGCTGATCTGCGATCGCCCCACGAGAGTGCATCCGGCTC 3004
QY 901 CTGATCGACCGCATCTCGGTTCATGTCGCGAGGTGCGCTCCAGSCCGGTGTGGCTGC 960
Db 3005 CTGTCGACAGGATCGAGCGGATGTCGAGAGTCCGCCCGCGCGAGTTCGAGGCC 3064
QY 961 GATGTGGATGAATTTTCAATCTCAGCCCCGTCGAGCTGGCTCTACCATGTTGGAGCGC 1020
Db 3065 GCGGTGAACGGGTTCCTTCGCGCTCAGCCCTGTGCGGTCTGTCTCCGTTGCTGATCGC 3124
QY 1021 GTTCGGGAAGCGGCTTCGCGCTTCAGTTTCACACACCGGGATATCAGCAGTGGGCGGC 1080
Db 3125 GTGCGCGACGCGCGCTTCGAACTCGGCTTCACCCATCGCGACATCACAGAGCGGCGAGG 3184
QY 1081 CACGACTCGATGTTTCATGCGCCAGGTTCACGAGCTCGGAATGTTTTCGTTCCAGCCGT 1140
Db 3185 CAGACTCGATGTTTCATGCGCCAGATCACCGAGCTCGGAATGTTTTCGTTCCAGCCGT 3244
QY 1141 GCTGCGGAGCGACGTTTCCGAAAGATGGAACCGATTTTCGATGACCTTCGCAAGGAACT 1200
Db 3245 GCGGGCGAAGCCATGTGCGGAGGAATGTCGATTCGAGATCTGCGGAAGGGGACG 3304
QY 1201 GAGGTTGCTTCGCGGTAATGAAGGCACTTGACCGGTAA 1239
Db 3305 GATGTGTCCTTCAGTCTGTCGCGCGCTTGACCGGTGA 3343
```

Search completed: May 4, 2004, 15:00:17
Job time : 5130 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 12:19:37 ; Search time 614 Seconds
(without alignments)
8572.504 Million cell updates/sec

Title: US-10-045-063-1
Perfect score: 1239
Sequence: 1 gtgacctgcagaagcgca.....tgaggcacttgaccggttaa 1239

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 3373863 seqs, 2124099041 residues
Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N Geneseq_29Jan04:*
1: Geneseqn1980s:*
2: Geneseqn1990s:*
3: Geneseqn2000s:*
4: Geneseqn2001as:*
5: Geneseqn2001bs:*
6: Geneseqn2002s:*
7: Geneseqn2003as:*
8: Geneseqn2003bs:*
9: Geneseqn2003cs:*
10: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1238	99.9	1239	5	AAS02276 Arthrobac
2	1238	99.9	1239	6	ABL41208 Arthrobac
3	739.8	59.7	1239	6	ABV72490 Nucleotid
4	739.8	59.7	1239	6	ABN86382 M. liquef
5	739.8	59.7	3343	6	ABV72491 Nucleotid
6	739.8	59.7	3343	6	ABN86383 Recombina
7	235.8	19.0	110000	6	ABAQ3041_05
8	220.6	17.8	110000	6	ABQ69245_05
9	220.6	17.8	110000	6	ABQ57197_04
10	182.2	14.7	1860	2	AAQ26728 Heat resi
11	179	14.4	1230	2	AAT39170 Bacillus
12	158.6	12.8	3880	6	ABQ71004 Listeria
13	158	12.8	7475	6	AAG15312 Plasmid p
14	150.2	12.1	1218	6	ABK73017 Bacillus
15	145.2	11.7	1450	2	AAX59333 Bacillus
16	122.6	9.9	32768	2	AAX13060 Enterococ
17	122.6	9.9	32768	6	ABS98855 Enterococ
18	121.2	9.8	110000	2	AAT42063_05
19	121.2	9.8	110000	2	AAT42063_06
20	111.6	9.0	582	6	ABK77524 Bacillus
21	102.8	8.3	30390	4	AAS59520 Propionib
22	102.8	8.3	30390	7	ACF64449 Propionib
23	88.6	7.2	1665	5	AAS92765 DNA encod

ALIGNMENTS

RESULT 1
AAS02276

ID AAS02276 standard; DNA; 1239 BP.

XX AAS02276;

XX AAS02276;

DT 18-JUL-2001 (first entry)

XX Arthrobacter aureusens carbamoylase (hyc) DNA.

DE Arthrobacter aureusens carbamoylase (hyc) DNA.

XX Hydatoninase; racemase; carbamoylase; whole cell catalyst; L-amino acid;

KW D-amino acid; 5-monosubstituted hydantoin; racemisation; hydrolysis; ds;

KW bacterial cell; Escherichia coli; fermentation; purification; hyc;

KW reproduction rate.

XX Arthrobacter aureusens.

OS Arthrobacter aureusens.

XX Key

FT Location/Qualifiers

FT 1..1239

FT /*tag= a

FT /product= "A. aureusens carbamoylase (hyc)"

XX WO200123582-A1.

XX 05-APR-2001.

XX 31-AUG-2000; 2000WO-EP008473.

XX 28-SEP-1999; 99US-00407062.

XX (DEGS) DEGUSSA-HUELS AG.

XX (UYST-) UNIV STUTTGART.

XX (HOFF) ROCHE DIAGNOSTICS GMBH.

XX Altenbuchner J, Mattes R, Syldatk C, Wiese A, Wilms B;

XX Bommarium A, Tischer W;

XX WPI; 2001-266169/27.

XX P-PSDB; AAU01557.

XX New whole cell catalyst for degrading hydantoin into amino acids,

XX comprises a hydantoinase, a racemase and a carbamoylase.

XX Example; Page 32-34; 51pp; English.

XX The sequence represents an Arthrobacter aureusens carbamoylase (hyc)

CC DNA. A whole cell catalyst for the conversion of 5-monosubstituted
CC hydantoins to L- or D-amino acids, comprises cloned genes encoding for a
CC hydantoinase, a racemase and a carbamoylase. These genes are
CC overexpressed in the cell according to their turnover rates. The
CC bioconversion consists of chemical and/or enzymatic racemisation of 5-
CC substituted hydantoins, ring opening hydrolysis achieved by a
CC hydantoinase, and hydrolysis of the N-carbamoyl amino acid produced by
CC hydantoinase to the amino acid by carbamoylase. Isolation of
CC microorganisms capable for stereospecific hydrolysis of hydantoins is
CC necessary due to the increasing demand for optically pure amino acids.
CC The cells are usually bacterial cells, e.g. from *Escherichia coli*, since
CC they have high reproduction rates and easy growing conditions. Since all
CC the enzymes are produced in one strain, there is a reduction of
CC fermentation and purification costs
XX

SQ Sequence 1239 BP; 234 A; 357 C; 406 G; 242 T; 0 U; 0 Other;

Query Match 99.9%; Score 1238; DB 5; Length 1239;
Best Local Similarity 100.0%; Pred. No. 1.1e-287;
Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 2 TGACCCCTGCAGAAAGCGCAGCGCGCATTCAGAAAGAGATCCGGGAGCTCTCCCGGT 61
DB 2 TGACCCCTGCAGAAAGCGCAGCGCGCATTCAGAAAGAGATCCGGGAGCTCTCCCGGT 61
QY 62 TCTCGGAGAACGCCCGGTGTTACCGGCTGACCTACACTCCAGAGCATGCCGCGCGC 121
DB 62 TCTCGGAGAACGCCCGGTGTTACCGGCTGACCTACACTCCAGAGCATGCCGCGCGC 121
QY 122 GGGAAACGCTATTCGGGCTATGAAAGCGCGCGCTTGAGCGTTCGTGAAGACGCACTCG 181
DB 122 GGGAAACGCTATTCGGGCTATGAAAGCGCGCGCTTGAGCGTTCGTGAAGACGCACTCG 181
QY 182 GAAACATCATCGGCCGACGTGAAGCAGCTGATCCGAGCTTCTCGCATCCGCGTGGTT 241
DB 182 GAAACATCATCGGCCGACGTGAAGCAGCTGATCCGAGCTTCTCGCATCCGCGTGGTT 241
QY 242 CACACTTCGATTCGTCCGAAACGGCGGATGTTTGATGGCACTGCAGCGTGGTGGG 301
DB 242 CACACTTCGATTCGTCCGAAACGGCGGATGTTTGATGGCACTGCAGCGTGGTGGG 301
QY 302 CCTTTAGGCTGCCCGGTGATCTGGAGAACGGCTACGTAATCGGCATCCATTTGAGT 361
DB 302 CCTTTAGGCTGCCCGGTGATCTGGAGAACGGCTACGTAATCGGCATCCATTTGAGT 361
QY 362 TCATCCGATCTGFGAGGAGGAAGGGCCCGTTTCAGCAGTGGCATGTTGGCGCGCGGG 421
DB 362 TCATCCGATCTGFGAGGAGGAAGGGCCCGTTTCAGCAGTGGCATGTTGGCGCGCGGG 421
QY 422 CCATTGACGGTTGTCGCGGACGGGAACCTGCACTCTTTGTTGATGAGGATGAGTGT 481
DB 422 CCATTGACGGTTGTCGCGGACGGGAACCTGCACTCTTTGTTGATGAGGATGAGTGT 481
QY 482 CCGTTAGGCGGCGCTACTGCTTCCGCTTGAAGCCGGCGGAACTGCAGGCTGCAGGCC 541
DB 482 CCGTTAGGCGGCGCTACTGCTTCCGCTTGAAGCCGGCGGAACTGCAGGCTGCAGGCC 541
QY 542 GCTCCGCGCGGACCTCGCTGCTTTTATCGAACTACATTAACAAAGGACCGATCTTCG 601
DB 542 GCTCCGCGCGGACCTCGCTGCTTTTATCGAACTACATTAACAAAGGACCGATCTTCG 601
QY 602 AGCAGGACCAATAGAGATCGAGTTGTAACCTCCATCTGTCGCTTCGCGCATTCGCGG 661
DB 602 AGCAGGACCAATAGAGATCGAGTTGTAACCTCCATCTGTCGCTTCGCGCATTCGCGG 661
QY 662 TTGCGCTCAAGCGAGAACGACCGCGGACCAACCCCATGCACTGCGCAGGATG 721
DB 662 TTGCGCTCAAGCGAGAACGACCGCGGACCGCGGACCAACCCCATGCACTGCGCAGGATG 721
QY 722 CGCTGGTACCGCGCTCTCATGGTGGAGGAGGTCAACCGGTTGTTCAACGAGATCGCG 781
DB 722 CGCTGGTACCGCGCTCTCATGGTGGAGGAGGTCAACCGGTTGTTCAACGAGATCGCG 781

QY 782 ATGCGACAGTGGCTACCGTTGGCCACCTCCACATGAGTGGCCCGCTGGAGGCAACAGGTCC 841
DB 782 ATGCGACAGTGGCTACCGTTGGCCACCTCCACATGAGTGGCCCGCTGGAGGCAACAGGTCC 841
QY 842 CGGGGAGGTGGACTTCACACTGGACCTGCGTTCTCCGCATGAGGAGTGGCTCCGCGTGC 901
DB 842 CGGGGAGGTGGACTTCACACTGGACCTGCGTTCTCCGCATGAGGAGTGGCTCCGCGTGC 901
QY 902 TGATCGACCGCATCTCGGTCACTGGCGGAGGTGCGCTCCAGGCGGTGCTGGTGGCGG 961
DB 902 TGATCGACCGCATCTCGGTCACTGGCGGAGGTGCGCTCCAGGCGGTGCTGGTGGCGG 961
QY 962 ATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGCTGGCTCTTACCATGCTGACGCGG 1021
DB 962 ATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGCTGGCTCTTACCATGCTGACGCGG 1021
QY 1022 TTCGGAAGCGGCTCGGCTTTCAGTTTCACACACCGGAGATATCAGAGTGGGCGGCGC 1081
DB 1022 TTCGGAAGCGGCTCGGCTTTCAGTTTCACACACCGGAGATATCAGAGTGGGCGGCGC 1081
QY 1082 ACGACTCGATGTTTCATCGCCCGAGTTCAGGACGTCGGAATGGTTTTCGTTCCAAAGCGGTG 1141
DB 1082 ACGACTCGATGTTTCATCGCCCGAGTTCAGGACGTCGGAATGGTTTTCGTTCCAAAGCGGTG 1141
QY 1142 CTGGCCGAGCCACGTTCCCGAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACTG 1201
DB 1142 CTGGCCGAGCCACGTTCCCGAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACTG 1201
QY 1202 AGGTTGCTCTCCGGGTAAATGAAGGACCTTGACCGGTAA 1239
DB 1202 AGGTTGCTCTCCGGGTAAATGAAGGACCTTGACCGGTAA 1239
RESULT 2
ABL41208
ID ABL41208 standard; DNA; 1239 BP.
XX
AC ABL41208;
XX
DT 07-MAY-2002 (first entry)
XX
DE Arthrobacter aureus polynucleotide SRQ ID NO 3.
XX
KW Allysine acetal; hydantoin; hydantoinase; carbamoylase; racemisation;
KW pharmaceutical; ds.
XX
OS Arthrobacter aureus.
XX
PN DE10037115-A1.
XX
PD 07-FEB-2002.
XX
PF 28-JUL-2000; 2000DE-01037115.
XX
PR 28-JUL-2000; 2000DE-01037115.
XX
PA (DEGS) DEGUSSA AG.
XX
PI Krimmer H, May O, Klement I, Drauz K, Reichert D;
XX
PS WPI; 2002-218024/28.
XX
PT Allysine acetal preparation in high yield, for use as pharmaceutical
PT intermediate, by reacting corresponding hydantoin with hydantoinase and D
PT - or L-specific carbamoylase.
XX
PS Disclosure; Page 5; 6pp; German.
XX
CC The invention relates to the preparation of allysine acetals (I) from
CC corresponding hydantoins (II) involving reaction of (II) with
CC hydantoinase and D- or L-specific carbamoylase, together with spontaneous
CC and/or enzyme-catalysed racemisation, where the enzymes are used in free,
CC immobilised or cell-enclosed form. The use of (I) obtained by the process

CC The specification also describes a hydantoinase enzyme. The enzymes are
CC used for production of proteins and optically-active amino acids, which
CC are applicable in synthesis of pharmaceuticals, chemicals or food
CC additives

XX	Sequence	1239 BP; 192 A; 387 C; 462 G; 198 T; 0 U; 0 Other;
	Query Match	59.7%; Score 739.8; DB 6; Length 1239;
	Best Local Similarity	74.8%; Pred. No. 6.1e-168;
	Matches	927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
Qy	1	GTGACCTTGCGAAAGCGAAGCGCGCGCATTTGAGAAAGAGATCCGGGAGCTCTCCCGG 60
Db	1	GTGACGCTGCGACGAGCGCGCGCGCATGCGATCGAGGAGGAGCTCTGGACTCTCTCCCGC 60
Qy	61	TTCTTCGGCAGAGGCCCGCGGTGTTACCGCGCTGACCTACACTCCAGAGCATGCCGCCGCG 120
Db	61	TTCTTCGGTCGAAAGGCCCGCGGTGACAGCTCTCAGTACACTCCGGAGACGCCGCCGCG 120
Qy	121	CGGGAACCGCTCATTTGCGGCTATGAAGCGCGCGCTTGAGCGTTCTGTGAAGACGCACTC 180
Db	121	CGAGAGTGATCTGTCCGCCCATGCGCGGATGAGCGTTCACGAGACGCTCTC 180
Qy	181	GGAAACATCATCGCGCGGAGCGTGAAGGCACTGATCCGGAGCTTCCTCGGATCGCGGTCCGT 240
Db	181	GGCAACATCATCTGGTCGGGTGAGGGAGCGACCCCGCTTCGCCGGCGATCGCCTTCGGC 240
Qy	241	TCACACTTCGATTCTGTCCTCGAAACCGCGGGATGTTTGATGGCACTCGAGCGGTGGTGTGC 300
Db	241	TGCGACTTCGACTCGGTCCGCAACCGCGGGATGTTTCGACGGCACCGCGGCGTGGTGTGC 300
Qy	301	GCCCTTGAGGCTGCCCGGCTGATCTCTGGAGNACGGCTACGTGATCGGCATCCATTGTGAG 360
Db	301	GCGCTCGAGGCTCGAGGGTGCTCTCGAGGAGCGGATATGTGAACCGTCACTCTCTCGAG 360
Qy	361	TTCATCCGATCTGTGAGGAGAAAGGGGCCCGCTTCAGCAGTGGCATGTTGGGCGCGCGG 420
Db	361	GTCAATCGCATCTGTCGAAGAGGAGGGCACCCGCTTCAGACGGCATGCTGGGCGGTGCG 420
Qy	421	GCCATTGCGGGTTGGTTCGCCGACAGGGAATGCACTCTTTGGTTGATGAGGATGGAGTG 480
Db	421	GCGATCCGGGGCTCTGTCCGACGCCGATCTGGACACCTCTGGTGGACGAGACGCGGTG 480
Qy	481	TCCGTTTAGGCAGCGCGCTACTGCTCTCGGCTTTGAAGCCGGGCGCAATGCAAGGCTCGAGCC 540
Db	481	ACGCTGCGGAGGCGGCCACCGGCTTCGGGCTTGAAACCGGTGAGCTCGGACGGCGGCC 540
Qy	541	CGCTCCGGCGGGAACCTGCTGCTTTTATCGAACTACACATTGAACAAAGACACGATCCTC 600
Db	541	CGTACGAGGGATGACCTTCGCGCTTCATCGAGTTGCACATCGACGAGGGGCGATCCTC 600
Qy	601	GAGCAGAGCAAAATAGAGATCGGAGTTGTAACTCCATCTGTTGGCTTTCGCGCATTCGGG 660
Db	601	GAGCAGAGAAAGTGGAGATCGGCTCGTGNACGGGGATCGTCTGGTGTTCGGGCTTCGGG 660
Qy	661	GTTGCCCTCAAGGCGAAGGCAACGACGCGCGGCAACACCCCATATGCACCTTCGCGCAGGAT 720
Db	661	ATCACGGTGGAGGCGAGGAGCGACACGCGGGGACGACCCCATATGCACCTTCGCGCAGGAC 720
Qy	721	GCGCTGTTACCCGCGCTCTCATGGTCGAGGAGGTCAACCGTTCGTCAACGAGATCGCC 780
Db	721	GCGCTGTTGCGGCGGCGCTCATGGTCGAGAGATCAATCGGTTCTCAACGAGATCGCG 780
Qy	781	GATGGCACAGTGGCTACCGTTGGCCACTCACTAGTGGCCCCCGGTGAGGCGAACAGGTC 840
Db	781	GACGGCACGGTGGCGACCGCTCGGCGCACTCACGGTGAACCCCTGGTGGGCTCAACAGGTT 840
Qy	841	CCGGGGAGTGGAATTCACACTGGACCTCGTTCTCCGATGAGGAGTTCGCTCCGCGTG 900
Db	841	CCCGGGGGCTCGAGTTTCAGCTCGATCTCGGATTCGCCCCACAGGAGGTGATCCGCGTC 900
Qy	901	CTGATCGACCGCATCTCCGTCATGGTCGGGAGGTTCGCTCCACAGGCGCGGTGGTGTGCC 960

Query Match 59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity 74.8%; Pred. No. 6.1e-168;
Matches 927; Conservative 0; Mismatches 312; Indels 0;

Qy	1	GTGACCTCGAGAAAGCGCAGCGCGCATCTGAGAAAGAGATCTCGGAGAGCTCTCCCGG	60
Db	1	GTACCGCTGAGAGCGCGGCGCCATCGCATCGAGAGAGCTCTGGAATCTCTCTCCCGG	60
Qy	61	TTCTCGGCAAGAGGCCCCCGGTGTTACCGGGCTGACCTACATCCAGAGCATGCGCGCGGG	120
Db	61	TTCTCGGTGAAAGGGCCCCGCGTGACAGTCTCAGGTACACTCCGGAGCACGCGCGCGG	120
Qy	121	CGGAAACGCTCATTTGCGGCTATGAAGCGCGCGCTTGAGCGTTCTGTGAAGACGCACTC	180
Db	121	CGAGAGTGATCGTCCGCGCCATGACAGCGGACGCGGGCTGAGCGTCCACAGAGACGCTCTC	180
Qy	181	GGAAACATCATCGGCGCGCTGAAGCGCACTGATCCGGAGTTCCTCGCATGCGGGTCTGT	240
Db	181	GGCAACATCATCGTTCGGGCTGAGGGAGCAACCCGCTCTGCCGCGATCGCTTCGCG	240
Qy	241	TCACACTTCGATTCTGTCGAAACCGCGGGATGTTTATGGCACTGCAGCGTGTGTGC	300
Db	241	TGCGACTTCGATCGTCTCGCAACCGCGGGATGTTTCAGCGGCACTGTTGGCGCGCGG	300
Qy	301	GCCCTTGAGGCTGCCCGGTGATGCTGAGAAACGGCTACGTGAATCGGCAATCAATTGAG	360
Db	301	GGCTCTCAGGCTTCGAGGGTCTGTCAGAGAGCGGATATGTGAACCGTCACTCTCTCSAG	360
Qy	361	TTCATCCGATCGTGAGAGGAAGGGCGCCGCTTTCAGCAGTGGCATGTTGGCGCGCGG	420
Db	361	GTATCCGATCGTCTGAGAGAGGGCACCCGCTTTCAGCAGCGCATGCTGGCGGTCTCG	420
Qy	421	GCAATTGACAGGTGGTTCGCGACAGGAACTGCACTCTTTGTTGATGAGATGAGAGTG	480
Db	421	GGCATCCGGGCTCGTGTCCGACGCGCATCTGGACACCCCTGGTGGACGAAGACGCGTG	480
Qy	481	TCCGTTAGCGAGCGGCTACTCCCTTTCGGCTTGAGCGGGCGCACTGCAGGCTCGACGC	540
Db	481	ACGGTCGGGAGCGGCGCACGCGCTTCGGGCTGTGAAACGGGTGAGCTGTGCGACGGCGG	540
Qy	541	CGCTCCGCGCGGACCTCGTGTCTTTATCGAACTTACACATTTGAACAAAGACCGATCTTC	600
Db	541	CGTACAGGGATGACCTTCGCGCTTCATCGAGTTGCATCATCGACAGGGCGCATCTCTC	600
Qy	601	GAGCAGGACAAATAGATCGGAGTGTAACTCTCAATCGTTGGCGTTTCGCGAATTCGGG	660
Db	601	GAGCAGGAAGGTGAGATCGGCTCGTGACGGGGATCGTCTGTGTCTCGCGCTTCGCGG	660
Qy	661	GTTCGCGTCAAGGCGAGAGCGACACGCGCGCACACCCCATGCACCTGCGCCAGGAT	720
Db	661	ATCACGGTGGAGGGCAGAGCGACACGCGCGGACGACCCCATGCACCTTCGCGGAGGAC	720
Qy	721	GGCTTGTTACCGCGCGCTCATGTGTGAGGAGGTCAACCGGTTCTGTCACAGATTCGCG	780
Db	721	GGCTGTGTCGCGCGCGCTCATGTGTGAGAGATCAATCGGTTCTGTCACAGAGATCGCG	780
Qy	781	GATGGCAGATGCTTACGTTGGCCACTCACAGTGGCGCCCCCGTGGAGGCAACAGGTC	840
Db	781	GACGGCAGGTGGCGACCGTCCGGCACCTCACGGTGAACCTTCGTTGGGCTCAACCAAGTT	840
Qy	841	CCGGGGGAGTGCATTCACACTGGACCTGGTTCTCCGCATGAGAGTCTGCTCCGCGGTG	900
Db	841	CCGGGGGCGTCGAGTTTACGGTCAATCTGCGATTCGCCCAAGAGAGTGCATTCGGGCTC	900
Qy	901	CTGATCGACCGCATCTCGGTTCATGTCGGCGAGGTTCGCTCCCGAGCGCGGTGTGCTGCC	960

RESULT 4

ABN86382

ID ABN86382 standard; DNA; 1239 BP.

AC ABN86382;

DT 21-OCT-2002 (first entry)

DE M. liquefaciens CHase encoding DNA.

AA Hydantoin racemase; HRase; bacterium; hydantoin; N-carbamoyl-amino acid;
KW
AA
KW pharmaceutical; chemical; food additive; hydantoinase; CHase; gene; ds.

OS *Microbacterium liquefaciens*.

XX	Key	Location/Qualifiers
FH	CDS	1. .1380
FT		

```

E1 E2 E3
/*tag= a
/*tag= b
/*tag= c

```

FT	/note= "N-carbamoyl
1	
2	
3	
4	
5	
6	
7	
8	
9	
10	
11	
12	
13	
14	
15	
16	
17	
18	
19	
20	
21	
22	
23	
24	
25	
26	
27	
28	
29	
30	
31	
32	
33	
34	
35	
36	
37	
38	
39	
40	
41	
42	
43	
44	
45	
46	
47	
48	
49	
50	
51	
52	
53	
54	
55	
56	
57	
58	
59	
60	
61	
62	
63	
64	
65	
66	
67	
68	
69	
70	
71	
72	
73	
74	
75	
76	
77	
78	
79	
80	
81	
82	
83	
84	
85	
86	
87	
88	
89	
90	
91	
92	
93	
94	
95	
96	
97	
98	
99	
100	

XX
PN
FD1188826-A2

XX

XX XX

PR 08-MAR-2001; 2001JP-00065815.

PA (AJIN) AJINOMOTO CO INC.

XXIX

XX

DR P-PSDB; ABB80945.

XX Now E-substituted hydro

PT active amino acid

FI Microbacterium species.
XX
PS Disclosure: Page 27-28: 40pp: English.

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

is derived from a Microbacterium by culture, disruption or lysis, and purification. HRAse has a high working temperature of 5-60 plusOC and is used to racemize optically active 5-substituted hydantoins for subsequent enzymatic conversion to N-carbamoyl-amino acids and then optically active amino acids (useful in pharmaceuticals, the chemical industry and as food additives). The present sequence represents a DNA encoding the M.

CC liquefaciens N-carbamoyl-amino acid-L-hydrolase (Chase) enzyme
XX Sequence 1239 BP; 192 A; 387 C; 462 G; 198 T; 0 U; 0 Other;
SQ
Query Match 59.7%; Score 739.8; DB 6; Length 1239;
Best Local Similarity 74.8%; Pred. No. 6.1e-168;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAAAGAGATCCGGAGCTCTCCCGG 60
DB |||||
QY 61 TTCTCGGAGAGGCCCCGGTGTACCGGCTGACCTACACTCCAGAGATCGCGCGG 120
DB |||||
QY 61 TTCTCGGAGAGGCCCCGGTGTACCGGCTGACCTACACTCCAGAGATCGCGCGG 120
DB |||||
QY 121 CGGGAACCGCTCATTTCCGGCTATGAAAGCGCGCTTCAGAGCTTCGTGAAGACGCACTC 180
DB |||||
QY 121 CGAGAGGTGATCGTCCGCCCATGACGCGACGGGCTGAGCGTCCAGAGGACGCTCTC 180
DB |||||
QY 181 GGAACATCATCGGCGGACGCTGAAGGCACTGATCCGGAGCTTCTCGCATCGCGTCTCGT 240
DB |||||
QY 181 GGAACATCATCGGCGGCTGAGGGGAGCGACCCCGCTCTGCGGCGATCGCTTCGGC 240
DB |||||
QY 241 TCACACTTCGATCTGTCCGAACCGCGGGATGTTTGAATGGCACTGCAAGCGTGTGTGC 300
DB |||||
QY 241 TCACACTTCGATCTGTCCGAACCGCGGGATGTTTGAATGGCACTGCAAGCGTGTGTGC 300
DB |||||
QY 301 GCCTTTAGGCTGCCCGGTGATGCTGGAGAACGGCTAGCTGAATCGGATCCATTTGAG 360
DB |||||
QY 301 GGGCTCGAGCTGCGGGTGTCTGAGGAGCGGATGTTGAACCGTATCTCTTCGAG 360
DB |||||
QY 361 TTTCATCGCGATCGTGGAGGAGGAGGCGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
DB |||||
QY 361 GTTCATCGCGATCGTGGAGGAGGAGGCGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
DB |||||
QY 421 GCATTCGAGGTTGTTGCCGACAGGAACTGGACTCTTTGGTGTGATGAGATGAGTGTG 480
DB |||||
QY 421 GCATTCGAGGTTGTTGCCGACAGGAACTGGACTCTTTGGTGTGATGAGATGAGTGTG 480
DB |||||
QY 481 TCCGTTAGGCGCGGCTACTGCTTCCGCTTGAAGCGGGGCACTGCGAGGCTCGAGC 540
DB |||||
QY 481 ACGTTCGCGAGCGGCGGCGGCTTCCGGCTTGAAGCGGGGAGTTCGGAGCGCGGCG 540
DB |||||
QY 541 CGCTCCGCGCGGACCTCGTGTCTTTATCGAACTACATTCGAACGAGGACCGATCTCTC 600
DB |||||
QY 541 CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCATCGAGCGGCGGATCTCTC 600
DB |||||
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACCTCATCTGTTGGGCTTCGCGCATTCGGG 660
DB |||||
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACCTCATCTGTTGGGCTTCGCGCATTCGGG 660
DB |||||
QY 661 GTTCCGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
DB |||||
QY 661 ATACGCTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 720
DB |||||
QY 721 GCGTGGTACCGCGGCTCTCATGTTGAGGAGGAGTCAACCGGTTCTGTCACAGAGATCGCC 780
DB |||||
QY 721 GCGTGGTACCGCGGCTCTCATGTTGAGGAGGAGTCAACCGGTTCTGTCACAGAGATCGCC 780
DB |||||
QY 781 GATGACAGTGGTACCGTTCGCACTTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB |||||
QY 781 GATGACAGTGGTACCGTTCGCACTTCAGTGGGCGGCGGCGGCGGCGGCGGCGGCGG 840
DB |||||
QY 841 CGGCGGAGGTGAGTTCACATGACCTGACCTGCTTCTCGCATGAGGAGTTCGCGCTG 900
DB |||||
QY 841 CGGCGGAGGTGAGTTCACATGACCTGACCTGCTTCTCGCATGAGGAGTTCGCGCTG 900
DB |||||
QY 901 CTGATCGAGCGGATCTGCTGATGTTGCGGAGGAGTTCGCGCTTCGCGAGGAGTTCGCG 960
DB |||||
QY 901 CTGATCGAGCGGATCTGCTGATGTTGCGGAGGAGTTCGCGCTTCGCGAGGAGTTCGCG 960
DB |||||
QY 961 GATGAGGATGAATTTTCAATCTCAGCGCGGTGCGAGTGGCTCTTACCATGTTGAGCGCC 1020

DB 961 GCGGTGAACGGGTCTCTTCGCGCTCAGCCCTGTCGGTCTGTCTCCGCTGCTGCGATCGC 1020
QY 1021 GTTTCGGAAGCGGCTTCGCGCTTCAGTTTCACACCGGATATCAGCAGTGGGCGGCG 1080
DB |||||
QY 1021 GTTTCGGAAGCGGCTTCGCGCTTCAGTTTCACACCGGATATCAGCAGTGGGCGGCG 1080
DB |||||
QY 1081 CACGACTCGATGTTTCATCGCCAGGTCACGAGCTCGGAATGGTGTTCGTTCCCAAGCGCT 1140
DB |||||
QY 1081 CACGACTCGATGTTTCATCGCCAGATCACGAGCTCGGAATGGTGTTCGTTCCCAAGCGCT 1140
DB |||||
QY 1141 GCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB |||||
QY 1141 GCTGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1200
DB |||||
QY 1201 GAGGTGTCCTCCGGTAAATGAAGCACTTGACCGGTAA 1239
DB |||||
QY 1201 GATGTTGTCCTTCACGTCGTGACGGGCTTGACCGGTGA 1239
DB |||||
RESULT 5
ABV72491
ID ABV72491 standard; DNA; 3343 BP.
XX
AC ABV72491;
XX
DT 29-JAN-2003 (first entry)
XX
DE Nucleotide sequence from Microbacterium liquefaciens.
XX
KW Hydatonase; enzyme; N-carbamyl-L-amino acid hydrolase;
KW optically-active amino acid; ss.
XX
OS Microbacterium liquefaciens.
XX
PN WO200272841-A1.
XX
PD 19-SEP-2002.
XX
PF 08-MAR-2002; 2002WO-JP002173.
XX
PR 08-MAR-2001; 2001JP-00065814.
PR 27-SEP-2001; 2001JP-00298619.
XX
PA (AJIN) AJINOMOTO CO INC.
XX
PI Takenaka Y, Suzuki S, Onishi N, Yokozeki K;
XX
WPI; 2002-691806/74.
XX
Microbacterium-originate DNAs encoding hydantoinase and N-carbamyl-L-
amino-acid hydrolase and transformant cells, useful for production of
proteins and optically-active amino acids, applicable in synthesis of
e.g. pharmaceuticals.
XX
PS Claim 27; Page 70-72; 83pp; Japanese.
XX
The present sequence is derived from Microbacterium liquefaciens. The
specification describes a hydantoinase enzyme and a N-carbamyl-L-amino
acid hydrolase enzyme. The enzymes are used for production of proteins
and optically-active amino acids, which are applicable in synthesis of
pharmaceuticals, chemicals or food additives
XX
SQ Sequence 3343 BP; 599 A; 1041 C; 1173 G; 530 T; 0 U; 0 Other;
Query Match 59.7%; Score 739.8; DB 6; Length 3343;
Best Local Similarity 74.8%; Pred. No. 7.7e-168;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAAAGAGATCCGGAGCTCTCCCGG 60
DB |||||
QY 2105 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAAAGAGATCCGGAGCTCTCCCGG 2164

QY 61 TTCTCGGAGAGGCGCCGGTGTACCGCGCTGACCTACATCCAGAGCATGCCGCGG 120
DB 2165 TTCTCGGTCAAGGGCCCGCGTGAACGTCTCACTGTAACATCCGAGGACGCGCGCG 2224
QY 121 CGGGAACGCTCATTTGCGGCTATGAAGCGCGCTTGAGCGTTGTTGAAGACGACATC 180
DB 2225 CGAGAGGTGATCGTCCGCCCATGACGCGGAGCGGGCTGAGCGTCCACGAGGACGCTCTC 2284
QY 181 GGAACATCATCGCGGAGTGAAGCACTGATCCGAGCTTCCTCGCATCGCGTCCGGT 240
DB 2285 GGCAACATCATCGGTGCGGTGAGGGAGCGACCCCGCTCTCCCGCGCATCGCTTCGCG 2344
QY 241 TCACACTTCGATCTCTCGAAACCGCGGATGTTTGAATGGCACTGCGAGCGTGTGTGC 300
DB 2345 TCGCACTTCGATCGGTTCGCAACCGCGGATGTTTCGACGGCACCGCGGCTGTGTGC 2404
QY 301 GCGCTTAGGCTGCCGCGGTGATGTGGAGAACGCGTACGTGAATCGGATCCATTGAG 360
DB 2405 GCGCTCGAGGCTCGAGGGTGTGTCAGAGAGCGGATGTGAACCGTCACTCTCTCGAG 2464
QY 361 TTTATCGGATCGTGGAGGAGGAGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
DB 2465 GTATTCGCGATCGTGAAGAGAGGAGCGACCCCGCTTCAGAGCGGCATGCTGGCGGTGCG 2524
QY 421 GCCATTGCAAGGTGTGTCGCCACAGGGAACCTGGACTCTTTGTTGATGAGGATGAGTG 480
DB 2525 GCGATCGCGGGCTGTGTCCGACCGCGATCTGGACACCTGTGTGACGAGACGCGGTG 2584
QY 481 TCCGTTAGCGAGCGGCTACTGCTTGGCTTTGAAGCGCGGCGAACTGCGAGGCTCGAGCC 540
DB 2585 ACGGTGCGGAGCGGCCACGCGCTTCGGGCTTGGAAACCGGCTGAGCTGCGAGCGCGCC 2644
QY 541 CGCTCGCGCGGACCTGCTGCTTTTACGAACCTACACATTTGAACAGGACCGATCTCTC 600
DB 2645 CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCACATCGAGCGGCGCGATCTCTC 2704
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACCTCATCGTTGGGCTTCGCGCATTTGCGG 660
DB 2705 GAGCAGGAGAGGTGAGATCGGCTGTGACGGGATCGTGTGTCGCGCCTTCGCG 2764
QY 661 GTTCCGCTAAAGGAGAGGACCAACGCGCGCAAAACCCCATGCACTGCGCCAGGAT 720
DB 2765 ATACGCTGGAGGAGGAGGACCAACGCGCGGACGACCCCATGCACTGCGCGAGGAC 2824
QY 721 GCGCTGTTACCGCGCTCTCATGTGTGAGGAGGTTCAACCGGTTGTTGTTCAACGAGATCGCC 780
DB 2825 GCGCTGTTGCGCGCGGCTCATGTGTGAGAGATCAATCGGTTGTTCAACGAGATCGCG 2884
QY 781 GATGCGACAGTGGCTACCGTTGCGCCACCTTCACAGTGGCGCCCGTGGAGCAACGAGTTC 840
DB 2885 GACGCGACCGTGGCGACCGTTCGCGCACCTTCACGTTGACCCCTGTTGGCTCAACACGTT 2944
QY 841 CCGGGGAGGTGGAATTCACTGACCTGTCGTTCTTCGCGCATGAGGAGTGTGCTCGCGTG 900
DB 2945 CCGCGGGGCGTTCAGTTCACTGCTCATGCTGATGCGATGCGCCCGCAGGAGTGTGATCGGCTC 3004
QY 901 CTGATCGACCGCATCTGCTCATGCTGCGGAGTGTGCTCTCCAGCGGCTGTGCTGCC 960
DB 3005 CTGTTGCGACGAGATCGAGCGCATGTGGGAGAGTGTGCGCGCGCGCGGAGTTCAGGCCC 3064
QY 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGCGAGTGGCTTCCTACCATGTGACGCGCC 1020
DB 3065 GCGGTGAACGGGTTCTTCGCGCTCAGCCCTGTGCTGTCTGCTCCGTTGTTGATCGC 3124
QY 1021 GTTCGCGAAGCGGCTTCGCGCTTCAGTTTCAACACCGGATATCAGCAGTGGGGCGGCG 1080
DB 3125 GTGCGCGAGCGCGCGTTCGGAATCTCGGCTTCACCCATCGCGCATCACGAGCGGGCAGGG 3184
QY 1081 CACGACTCGATGTTTATATCCCGAGGTTCAGGACGTGCGAATGTTTTCGTTCCAGCCGT 1140
DB 3185 CACGACTCGATGTTTATGTCGCCAGATCACCGACGTGCGAATGTTTTCGTTCCCGCGCG 3244
QY 1141 GCTGCGCGGAGCGACGTTTCCGGAAGATGGACCGATTTTCGATGACCTTTCGCAAGGAACT 1200

DB 3245 GCCGGCGAAGCCATGTGCGGAGGAATGTCCTGATTCGAGCATTCGCGAAGGGACG 3304
QY 1201 GAGTTGTCTCCGGGTAAATGAAGCACTTGACCGGTAA 1239
DB 3305 GATGTGGTCTTCACGTCGTGACGCGCTTGACCGGTGA 3343
RESULT 6
ABN86383
ID ABN86383 standard; DNA; 3343 BP.
XX AC ABN86383;
XX 21-OCT-2002 (first entry)
XX Recombinant DNA encoding proteins involved in L-amino acid production.
DE Hydatonin racemase; HRase; bacterium; hydatonin; N-carbamoyl-amino acid;
KW pharmaceutical; food additive; hydatoninase; HRase; Chase; gene; ds.
XX Microbacterium liquefaciens.
OS
FH Key Location/Qualifiers
FT CDS 1..711
FT /*tag= a
FT /product= "HRase"
FT /note= "hydatonin racemase"
FT 729..2108
FT /*tag= b
FT /product= "HRase"
FT /note= "hydatoninase"
FT 2105..3343
FT /*tag= c
FT /product= "CHase"
FT /note= "N-carbamoyl-amino acid-L-hydrolase"
XX EP1188826-A2.
PN 20-MAR-2002.
XX 09-AUG-2001; 2001EP-00119275.
XX 13-SEP-2000; 2000JP-00278571.
PR 08-MAR-2001; 2001JP-00065815.
XX (AJIN) AJINOMOTO CO INC.
XX Suzuki S, Onishi N, Yokozeki K;
XX WPI; 2002-510588/55.
DR P-PSDB; ABB80943, ABB80944, ABB80945.
XX New 5-substituted hydatonin racemase, useful in production of optically active amino acids, comprises high working temperature, from Microbacterium species.
PT Disclosure; Page 30-31; 40pp; English.
XX The invention relates to a 5-substituted hydatonin racemase (HRase) that is derived from a Microbacterium by culture, disruption or lysis, and purification. HRase has a high working temperature of 5-60 plusOC and is used to racemize optically active 5-substituted hydatonins for subsequent enzymatic conversion to N-carbamoyl-amino acids and then optically active amino acids (useful in pharmaceuticals, the chemical industry and as food additives). The present sequence represents a M. liquefaciens recombinant DNA containing structural genes encoding proteins such as HRase, HRase and Chase involved in L-amino acid production
XX Sequence 3343 BP; 599 A; 1041 C; 1173 G; 530 T; 0 U; 0 Other;
Query Match 59.7%; Score 739.8; DB 6; Length 3343;
Best Local Similarity 74.8%; Pred. No. 7.7e-168;

Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;			
Qy	1	GTACCTCTCAGAAAGCGAAGCGCGCGCAITTGAGAAAGAGATCGGGAGCTCTCCCGG	60
Db	2105	GTGACGCTGAGCAGCGCGGCGGATCGCATCGAGGAGGAGCTCTGGACTCTCTCCCGC	2164
Qy	61	TTCTCGGCAGAGGCCCCCGGTGTACCGGCTGACCTACACTCAGAGCATGCGCGCGG	120
Db	2165	TTCTCGGTGGAAGGGCCCCGCGTGACACGCTCTCACGTACACTCCGAGCAGCGCGCGG	2224
Qy	121	CGGAAACGCTCATTTGGCGGTATGAAAGCGCGCTTGGAGCGTTCGTGAAGACGCACTC	180
Db	2225	CGAGAGGTGATCGTCCGCGCATCGAGGACGGGCTGAGCGTTCACGAGGACGCTCTC	2284
Qy	181	GGAAACATCATCGGCCGACGTGAAGGCACTGATCCGGAGCTTCTCGCATCGCGTCTCGT	240
Db	2285	GGCAACATCATCGTCCGGGTGAGGGAGCGACCCCGCTCTGCGCGGATCGCTTCGCG	2344
Qy	241	TCACACTTCGATTCTGTCCGAACCGGGGATGTTTGTATGGGCACTCGAGCGTGGTGTGC	300
Db	2345	TGCACTTTCGACTCGGTCCGCAACCGGCGGATGTTCCAGCGGCACCGCGGCGTGTGTGC	2404
Qy	301	GCCTTTGAGGCTGCCCGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTTGAG	360
Db	2405	GCCTTCGAGGCTCGAGGGTCTCGAGGAGCGGATATGTGAACGTCATCTCTCGAG	2464
Qy	361	TTCATCCGATCGTGAGGAGGAAGGGGCCGCTTCAGCAGTGGCATGTTGGCGGCCCGG	420
Db	2465	GTCACTCGCATCGTTCGAAGAGGAGGACCCCGCTTCAGCAGCGGCATGCTGGGCGTCTGC	2524
Qy	421	GCATTGCAAGGTTGTCGCCGACAGGGAACCTGGACTCTTTGTTGATGAGGATGAGTGTG	480
Db	2525	GGATCGCGGGGCTCGTGTCCGACGCGCATCTGGACACCCCTGTTGGACGAACGCGGTG	2584
Qy	481	TCGGTTAGGCGAGCGGCTACTGCTTCCGCTTGAAGCGGGCGAACTGCAAGGCTGCAGCC	540
Db	2585	ACGGTGGGAGCGCGCCACCGCTTCCGGCTGGAAACCGGCTGAGTTCGCGACGGCGGCC	2644
Qy	541	CGTCCGCGCGGACCTCGTGTGTTTTATCGAACTACATCTGAACCAAGACCGATCTCTC	600
Db	2645	CGTACGAGGATGACCTTCGCGCTTCATCGAGTTGCACATCGAGCAGGGCGGATCTCTC	2704
Qy	601	GAGCAGGAGCAATAGAGATCGAGTTGTAACTCTCATCTGTTGGCGTTTCGCGATTCGCG	660
Db	2705	GAGCAGGAGAAAGTGAGATCGCGTCTGTCAGCGGGATCGTCTGCTGCGCGCTTCCCG	2764
Qy	661	GTTCGCGTCAAGGCGAGAGCGACACCGCGGACAAACCCCATGCACTGCGCCAGGAT	720
Db	2765	ATCACGGTGGAGGCGAGGAGCGACCGCGCGGACCGACCCCATGCACTTGGCGGAGGAC	2824
Qy	721	GGCTGGTACCCCGCGCTCTCATGTGTAGGAGGTCAACCGGTTCTGTCACAGAGATCGGC	780
Db	2825	GGCTGTGTCCGCGCGGCTCATGTTGCGAGAGATCAATCGGTTCTCAACGNGATCGG	2884
Qy	781	GATGGCAGATGCTACCGTTGGCCACTCAGTAGTGGCCCCCGGTGGAGCAACAGGTC	840
Db	2885	GACGCGACGTTGGCGACCGTCCGCCACCTCAGCGTGACCCCTGTTGGGCTCAACAGGTT	2944
Qy	841	CCGGGGGAGTGGACTTCACATGGACCTGCGTTCTCCGATGAGGAGTTCGCTCCGCGT	900
Db	2945	CCGGGGGCGTTCAGTTTCACGCTTCGATCTGCGATCCGCCACGAGGAGTTCGATCCGGCTC	3004
Qy	901	CTGATCGACCGCATCTCGGTTCAGTGTGCGGAGGTGCGCTCCAGCGCGGTGTGGCTGCC	960
Db	3005	CTGGTCGACAGGATCAGGCGATGGTGGCAGAAGTCCGCGCGCGCGGATCGAGGCC	3064
Qy	961	GATGTGGATGAATTTTCAATCTCAGCCCGGTGCACTGGCTCTTACCATGTGTGACGCC	1020
Db	3065	GGCGTGAACGGGTTCTTCGCGCTCAGCCCTGTCTGCTCTGCTCGGTGTGTGGATCGC	3124
Qy	1021	GTTCGCGAAGCGGCTCGGCTTGCAGTTTCACACCGCGGATATCAGCAGTGGGGCGGC	1080
Db	3125	GTGCGCGACGCGGCGCTCCGAACCTCGGCTTTCACCCATCGCGACATCAGCGCGGGCAGG	3184

Qy	1081	CACGACTCGATGTTTCATCGCCAGGTCACGAGCTCGGAATGTTTTGTTTCAAGCCGCT	1140
Db	3185	CACGACTCGATGTTTCATCGCCAGATCACGAGCTCGGAATGTTGTTCCGCCAGCCGC	3244
Qy	1141	GCTGGCGGAGGACGCTTCCGGAAGATGACCGGATTTGATGACCTTCGCAAGGAAC	1200
Db	3245	GCGGCGGAAGCATGTGCGGAGGAATGTTCCGATTTCCAGCATCTCGGAAGGGGAGC	3304
Qy	1201	GAGGTTGTCCTCCGGTAATGAAGGCACTTGACCGGTAA	1239
Db	3305	GATGTGTCCTTCACGTCGTGACGGCCTTGACCGGTGA	3343

RESULT 7

ABAO3041_05
Continuation (6 of 30) of ABAO3041 from base 500001 (Listeria monocytogenes EGD-e genom
WP Sequence split into 30 fragments LOCUS ABAO3041 Accession ABAO3041

WP	Fragment Name	Begin	End
WP	ABAO3041_00	1	110000
WP	ABAO3041_01	100001	210000
WP	ABAO3041_02	200001	310000
WP	ABAO3041_03	300001	410000
WP	ABAO3041_04	400001	510000
WP	ABAO3041_05	500001	610000
WP	ABAO3041_06	600001	710000
WP	ABAO3041_07	700001	810000
WP	ABAO3041_08	800001	910000
WP	ABAO3041_09	900001	1010000
WP	ABAO3041_10	1000001	1110000
WP	ABAO3041_11	1100001	1210000
WP	ABAO3041_12	1200001	1310000
WP	ABAO3041_13	1300001	1410000
WP	ABAO3041_14	1400001	1510000
WP	ABAO3041_15	1500001	1610000
WP	ABAO3041_16	1600001	1710000
WP	ABAO3041_17	1700001	1810000
WP	ABAO3041_18	1800001	1910000
WP	ABAO3041_19	1900001	2010000
WP	ABAO3041_20	2000001	2110000
WP	ABAO3041_21	2100001	2210000
WP	ABAO3041_22	2200001	2310000
WP	ABAO3041_23	2300001	2410000
WP	ABAO3041_24	2400001	2510000
WP	ABAO3041_25	2500001	2610000
WP	ABAO3041_26	2600001	2710000
WP	ABAO3041_27	2700001	2810000
WP	ABAO3041_28	2800001	2910000
WP	ABAO3041_29	2900001	2944528

Query Match 19.0%; Score 235.8; DB 6; Length 110000;
Best Local Similarity 50.9%; Pred. No. 2.4e-46;
Matches 586; Conservative 0; Mismatches 562; Indels 3; Gaps 1;

Qy	73	GCCCGCGGTGTACCGGCTGACCTACACTCCAGAGCATGCCCGCGCGGGAACGCTC	132
Db	75232	CGTCAGGCGACAACCTCGGCTTACATATAGCAAGAAGACCTCGACGCGGCAACTATTTA	75291
Qy	133	ATTGCGGCTATGAAGCGCGCCCTTACCGGATTCCTGCGATCGCGGTCGGTTTACACTTC	192
Db	75292	AAACAAGAAATCGCTTAAAGTGTGCTTACTGTCTTTCAGAAATGCGGATGCGGATATTTAC	75351
Qy	193	GCCGACGCTGAAGGACCTGATCCGAGCTTCTGCGATCGCGGTCGGTTTACACTTCGAT	252
Db	75352	GGACGACTAGAAGGGGAAGTCCAGATTTACAGCAGTAAATCGTGGTTCTCATTTTGAC	75411
Qy	253	TCTGTCCGAAACCGCGGATGTTTGTATGGCACTGCGAGCGTGTGTGCGCCCTTGAGGCT	312
Db	75412	TCTGTGCCAAATGTTGGTGTCTTTTACGCGACCGCTGCGTTCATTACCGGCTTGAAGTA	75471
Qy	313	GCCCGGCTGATGCTCGTGAGAACCGCTACGTTGAATTCGGATCCATTTGAGTTCATCGCGATC	372
Db	75472	GCCAGTGTTTTCCACGAAACAAATTAACCAACATTTTCCGCTTGAATTTATTTGCGATG	75531

QY	373	GTGGAGGAGGAGGGCCCGCTTCAGCAGTGGCATGTTGGGCGCGCGGCCCAATTCAGGG	432
Db	75532	GTGGAAGAAGAAGGCTCTCGTTTTGGCGCAGGACTTCTTGCTCTCAGCAGCAATCAGGC	75591
QY	433	TTGGTGGCGGCACAGGGAACTGGACTCTTTGGTTGATGAGGATGAGGTGTCGGTTAGGCAG	492
Db	75592	AAAGTCACAAACAGAAATGCTACAGAAATGAAAGATATAGATGGCATTACAGCTACGGAT	75651
QY	493	GGGGCTATGCGCTTGGCTTGAAGCGGGCGGAACTGCGAGGCTGCGAGCCGCTCCGGGGG	552
Db	75652	GCCATGGCGGCAACTAGGATTTAAACGCAACCGAGTACATAGAGCCATTGCGCAAAAGAG	75711
QY	553	GACCTCGCTCTTTATCAACTTACACATTTGACAGGACCGATCCTCGAGCAGAGCA	612
Db	75712	TCCGTGAAGACATTTATCGAACTGCAATCGAACAGGGCCCATCTCTGAAAATGCCAGT	75771
QY	613	ATAGAGATCGGAGTTGTAACTCCATCGTTGGCGTTTCGGGCATTGGCGTTCCCGTCAAA	672
Db	75772	GAAGCGTAGCGTAGTCATACAGTAGTTCGTTTAAACAATCAAAGTAAACCGTAAAA	75831
QY	673	GGCAGAGCGACACCGCGGCAACCCCATGACCTGCGCCAGATGCGCTGGTATGCC	732
Db	75832	GGGCAAGCGCGCATGCAAGCAACCCCGATGCTTGACCGAAAAGATGCGCTAGTTCC	75891
QY	733	GGCGCTCTCATGTTGAGGAGGTCAACCGTTGCTCAACGAGATCGCGATGCGCACAGTG	792
Db	75892	GCTGTGCAAAATTTTAGGCAATTTGCCAGAACTCGGTATCCAAGAGGTGGCGGAATGTG	75951
QY	793	GCTACGTTGGCCACTCACAGTGGCCCCCGTGGAGGCAACAGGTCCCGGGGAGGTG	852
Db	75952	CTAACAGTCGGCAACTAAACGTCTATCCAATGGCGCAATGTTATTCAGATAAGTC	76011
QY	853	GACTTCACACTGGACTCGGTTCTCGCATGAGGAGTCGCTCCGGTGTGATCGACCGC	912
Db	76012	ATTTTTCACCGTGGCAATTGCG---CGCAAAAGACGAAATTCAGTTTCAAAATACATTAGCA	76068
QY	913	ATCTCGGTATGTTGGCGAGTGCCTCCAGGCGGTGTCGCTGCCGATGTGATGAA	972
Db	76069	AAACGAAAGAAATCATTACAGCCACCGAAAATAATGTTACTTTCGAAAATAAAGAT	76128
QY	973	TTTTTCAATCTCAGCGCGGTGAGTGGCTCTCATCATGTTGGAGCGCGTTCGCGAAGCG	1032
Db	76129	ATGATTTACCAACAACCAACCCATTATCAAAAGAGATTTCATCAAGCTCTAACCGAAAGT	76188
QY	1033	GCTTCGGCTTGGAGTTTCAACACCGGGATATCAGCAGTGGGGCGGGGCCACGACTCGATG	1092
Db	76189	GCCGACCAATTGGGCTTTAAGTATCGAACAAATGTTAGCGGGCGGGCAGCAGCGATG	76248
QY	1093	TTTCATCGCCAGGTCAAGCACTCGGAATGTTTTGTTTCCAAAGCGGTGCTGGCGGAGC	1152
Db	76249	ATTTTCGCTAGTTTAAACGAAAGTGGCTTGAATTTTGTTCCTAGCCACACGCGATAAGT	76308
QY	1153	CACGTTTCCGAAAGATGGACCGATTTCGATGACCTTTCGCAAGGAACATGAGTTGTCCTC	1212
Db	76309	CATGCTCCAGAGATGGACCGATTACGACAGCTACAAAAGGAATTGAAGTCGTACTC	76368
QY	1213	CGGGTAATGAA	1223
Db	76369	AAAGACAGTAAA	76379

RESULT 8			
ABQ69245_05			
Continuation (6 of 31) of ABQ69245 from base 500001 (Listeria innocua DNA sequence #684.			
WP Sequence split into 31 fragments LOCUS ABQ69245 Accession Abq69245			
WP	Fragment Name	Begin	End
WP	ABQ69245_00	1	110000
WP	ABQ69245_01	100001	210000
WP	ABQ69245_02	200001	310000
WP	ABQ69245_03	300001	410000
WP	ABQ69245_04	400001	510000
WP	ABQ69245_05	500001	610000

WP	ABQ69245_06	600001	710000
WP	ABQ69245_07	700001	810000
WP	ABQ69245_08	800001	910000
WP	ABQ69245_09	900001	1010000
WP	ABQ69245_10	1000001	1110000
WP	ABQ69245_11	1100001	1210000
WP	ABQ69245_12	1200001	1310000
WP	ABQ69245_13	1300001	1410000
WP	ABQ69245_14	1400001	1510000
WP	ABQ69245_15	1500001	1610000
WP	ABQ69245_16	1600001	1710000
WP	ABQ69245_17	1700001	1810000
WP	ABQ69245_18	1800001	1910000
WP	ABQ69245_19	1900001	2010000
WP	ABQ69245_20	2000001	2110000
WP	ABQ69245_21	2100001	2210000
WP	ABQ69245_22	2200001	2310000
WP	ABQ69245_23	2300001	2410000
WP	ABQ69245_24	2400001	2510000
WP	ABQ69245_25	2500001	2610000
WP	ABQ69245_26	2600001	2710000
WP	ABQ69245_27	2700001	2810000
WP	ABQ69245_28	2800001	2910000
WP	ABQ69245_29	2900001	3010000
WP	ABQ69245_30	3000001	3011208

Query Match

17.8%; Score 220.6; DB 6; Length 110000;

Best Local Similarity 50.2%; Pred. No. 1.le-42;

Matches 572; Conservative 0; Mismatches 564; Indels 3; Gaps 1;

QY	73	GGCCCGGTGTACCCGGCTGACCTACACTCCAGAGCATGCCGCGCGCGGAAACGCTC	132
Db	68857	GGTCAGGAAACAATCGGCTTACATATAGCAAGAGACCTCGGCGCGGTAATTATTA	68916
QY	133	ATTGGGCTATGAAGCGGCGCCTTGAGCGTTTCTGAGAGCGCACTCGGAAACATCATC	192
Db	68917	AAAGAAGAAATGGCTTAAAGTAGGCTTACTGTCTCAGAAGATGCGATTGGAATATCTAC	68976
QY	193	GGCCGAGTGAAGCACTGATCCGAGGTTCTCGGATCGGCTCGGTTACACTTCGAT	252
Db	68977	GGACGACTAGAAGGCGCAATCCAGACATACCAGCAGTAATCGTCGGTTCTCATTTTGAC	69036
QY	253	TCTGTCCGAAACGGCGGATGTTTGATGGCACTGCAGCGTGGTGTGCGCCTTGAGGCT	312
Db	69037	TCTGTCCAATGTGTGCTTTTGACGCTCCAGCAGGTGTTATAACTGGGCTTGAAGTA	69096
QY	313	CCCCGGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTTGAGTTTCAATCGCATC	372
Db	69097	GCTAGCGTTTTTCATGAACAGCAATAAAACCCCAATTTTCCACTTGAATCATCGCAATG	69156
QY	373	GTGGAGGAGNAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGGCGCATTTGCAGGG	432
Db	69157	GTGGAAGAAGAGGATCTCGTTTCGGAGCGGGACTACTTTGCCTCCCGAACAAATTACAGGA	69216
QY	433	TTGTTGCGCGACAGGAACTGGAATCTTTGTTGATGAGGATGGAGTGTTCGTTAGGCAG	492
Db	69217	AAAGTTACAAAGAAATGCTACATGAATGAAGATATAAATGTTGTTACTGCCGAGAA	69276
QY	493	CGCGCTACTGCTTTCGGCTTGAAGCGGGCGAACTGCAGGCTGCAGCGCTCCGCGCGG	552
Db	69277	GCTATGGCAACCTTAGGATTTGACGCAAAATCAAGTTTCATACAGGCATTTCCGGTCCCAAGAT	69336
QY	553	GACCTCGCTGCTTTTATCGAACTACATTAAGCAAGACCGATCCTCGACAGCAGCA	612
Db	69337	TCCGTCAAGCTTTTATGTAACATACATCGAAACAGGCCCCAGTCCCTAGAAAAATGCGAAT	69396
QY	613	ATAGAGATCGGAGTTGTAAACCTCCATCGTTGGCGTTTCGCGCATTTGCGGTTTCCGCTCAA	672
Db	69397	GAAATGTTGCGCTAGTTGACACGGTCTGTGGTTTAAACAGAAATAAAAGTTTACCGTAAG	69456
QY	673	GGCAGAGCGACACCGCGGCAAAACCCCATGACCTGCGCCAGGATGCGCTGGTACCC	732
Db	69457	GGTCAAGCAGGTACGCGCGGAATACCCCAATGCTAGAACGAAAAGATGCTCTAAGTGCA	69516

```
QY 733 GCCGCTCTCATGTGAGGAGGTCAACCGGTTTCTGTCACACAGATCGCCGATGCGACAGTG 792
|||
Db 69517 GCTGTCAAATCTTAATAAGTCTCTGAACCTAGCTATCCAAAGCGCGGAACTGTC 69576
|||
QY 793 GCTACCGTTGGCCACCTCACAGTGGCCCCGGTGGAGGCAACAGGTCCCGGGGAGGTG 852
|||
Db 69577 TTAACAAATTGGCAAACTTAATGTCTACCCGAACGGCGCTAATGTAATACCAAAACAAAGTC 69636
|||
QY 853 GACTTCACACTGACCTGCGCTTCTCCGATGAGGAGTCGCTCCGCGTGTGATCGACCGC 912
|||
Db 69637 GTTTTCACAGTAGATATTCGAGCAAAAGAGCAATCCACGTCCAAATACATTTAGAAAA 69696
|||
QY 913 ATCTCGGTGATGTGCGGAGGTGCGCTCCAGCGCGGTGTGCTGCCGATGTGGATGAA 972
|||
Db 69697 A---CAAAAAGATTATTAAGCTTTTCGAAAAAAGCGGTATTATGTGTGAATAGAGAT 69753
|||
QY 973 TTTTTCAAATCTACGCGCGGTGACGTGGCTCCTACCATGTGTGACGCGGTTCCGAGCG 1032
|||
Db 69754 ATGCTCTATGAAAAAACAACACATTTATCAACAGAAATTCATCAAGCTTTGACCGAAAGC 69813
|||
QY 1033 GCCTCGGCTTCGAGTTTCAACACCGGGATATCAGCAGTGGCGCGGCCACGACTCGATG 1092
|||
Db 69814 GCGGATAAACTTGGTCTTAATATCGAACATGTTAGCGGCGCAGGACACGATGCAATG 69873
|||
QY 1093 TTCAATCGCCAGGTCAAGGAGTGGGAATGTTTTCGTTTCCAAAGCCGTGCTGCCGAGC 1152
|||
Db 69874 ATTTTCGAAAGTTTAAACCGAAGTAGGACTTATTTTGTACCAAGCCACAAAGGTATAAGC 69933
|||
QY 1153 CAGGTTCCGGAAGATGGACCGATTTTCGATGACCTTCGCAAGCACTGAGGTGTCCT 1211
|||
Db 69934 CATGCACCAAGAAATGGACCGATTACGACAACTCCAAAAGGATCGAAGTCGTACT 69992
|||
```

RESULT 9

AB067197_04
Continuation (5 of 12) of AB067197 from base 40001 (Listeria innocua contig DNA sequenc
WP Sequence split into 12 fragments LOCUS AB067197 Accession Abq67197

WP	Fragment Name	Begin	End
WP	AB067197_00	1	110000
WP	AB067197_01	100001	210000
WP	AB067197_02	200001	310000
WP	AB067197_03	300001	410000
WP	AB067197_04	400001	510000
WP	AB067197_05	500001	610000
WP	AB067197_06	600001	710000
WP	AB067197_07	700001	810000
WP	AB067197_08	800001	910000
WP	AB067197_09	900001	1010000
WP	AB067197_10	1000001	1110000
WP	AB067197_11	1100001	1163020

Query Match 17.8%; Score 220.6; DB 6; Length 110000;

Best Local Similarity 50.2%; Pred. No. 1.1e-42; Mismatches 564; Indels 3; Gaps 1;

Matches 572; Conservative 0;

```
QY 73 GCGCCCGGTGTTACCGGCTGACCTACCTCCAGAGCATGCGCCCGCGGGAACGCTC 132
|||
Db 78591 GGTGAGGAACAACTCGGCTTACATATAGCAAGAGAGACTCGCGCGGCTAATTATTTA 78650
|||
```

```
QY 133 ATTCGCGGTATGAAACGCGCGCTTTGAGCGTTGCGTGAAGACGCACTCGGAAACATCATC 192
|||
Db 78651 AAGAAGAAATGGCTAAAGTAGGCTTACTGTCTCAGAGATGCGATTTGGAATATCTAC 78710
|||
```

```
QY 193 GCGCAGCTGAAAGCACTGATCGGAGCTTCTGCGCATCGCGTCCGTTTACACTTCGAT 252
|||
Db 78711 GGACGACTAGAGCGGCAATCCAGACATACCAGCAGTAATCGTCGGTTCTCATTTTGAC 78770
|||
```

```
QY 253 TCTGTCCGAAACCGCGGGATGTTTGATGGCACTGCGAGCGGTGTGTGCGCCCTTGAGGCT 312
|||
Db 78771 TCTGTCCGAAATGGTGTGCTTTTGACGTCGACGAGGTGTTAATACTGGCTTGAGTA 78830
|||
```

```
QY 313 GCCCGGCTGATGCTGAGAACCGGCTACGTGAATCGGCATCCCAATTTGAGTTTCATCGCGATC 372
|||
```

```
Db 78831 GCTAGCGTTTTCATGAACAGCAAAATAAAACCCCACTTTTCCACTTGAATATCATCGCAATG 78890
|||
QY 373 GTGGAGGAGGAAGGGCCCGCTTTCAGCAGTGGCATGTTGGCGCGCGGCCCATTTGAGGG 432
|||
Db 78891 GTGGAAGAAGAGGATCTCGTTTCGAGCGGGACTACTTGCCTCCCGCAACAATACAGGA 78950
|||
QY 433 TTGTCGCCGACAGGAACTGGACTCTTTGGTTGATGAGGATGGAGTGTCCGTTAGGCAG 492
|||
Db 78951 AAGATTTACAAAGAAATGCTACATGAATATAAAGATATAAATGGTGTATTCTGCCGCGAG 79010
|||
QY 493 CGGCGTACTGCGCTTCGGCTTGAAGCCGGGGAACCTGACGCTGACGCGCTCCGCGCG 552
|||
Db 79011 GCTATGGCAAACTTAGGATTTGACGCAAAATCAAGTTTATACAGCCCAATTCGGTCCAAGAT 79070
|||
QY 553 GACTCGGTGCTTTTATCGAACTTACACATTTGAACAGGACCGGATCCTCGAGCAGGAGCAA 612
|||
Db 79071 TCGGTCAAAGCTTTTATGAACTTACACATCGAACAGGCCAGCTCTAGNAATGCAAT 79130
|||
QY 613 ATAGAGATCGGAGTTGTAACTCCATCGTTGGCGTTTCGCGCATTTGCGGTTGCCGTCAAA 672
|||
Db 79131 GAAGATGTTGCGCTAGTTGACACGCTCGTTGTTTAAACAGAAATAAAAGTTTACCGTAAAG 79190
|||
QY 673 GGCAGAGGACCAACGCGCGGACACCCCACTGCATGCACCGCCAGGATGGCTGGTACCC 732
|||
Db 79191 GGTCAAGCAGGTACGCGCGGAACTACCCCAATGCTAGAACGAAAGATGCTCTAAGTGCA 79250
|||
QY 733 GCGGCTCTCATGTTGAGGAGGTCAACCGGTTCTGCAACGAGATCGCCGATGCGACAGTG 792
|||
Db 79251 GCTGTCAAATTTCTTAATAGCTTCTGAACTAGCTATCCAAGAGCGCGGCACTGTC 79310
|||
QY 793 GTTACCAATTTGGCAAACTTAATGTTCTACCCGACGCGCTAATGTTATACCAACAAGTC 79370
|||
QY 853 GACTTCACACTGACCTGCGTTCTCCGATGAGGAGTGCCTCGCGCTGCTGATCGACCGC 912
|||
Db 79371 GTTTTCACAGTAGATATTCGAGCAAAAGACGAAATCCACGTCCAAATAATACATTAGAAAA 79430
|||
QY 913 ATCTCGGTGATGTCGCGGAGTTCGCTCCAGCGCGGTGTCGCTCCGATGTTGAGTGA 972
|||
Db 79431 A---CAAAAAGATTATTCAGCTTTCGAAAAAAGCGGTATTATGTTGAAATAGAGAT 79487
|||
QY 973 TTTTTCATCTCAGCCCGGTGCGTCTCTACCATGTTGAGACCGCTTCGCGAAGCG 1032
|||
Db 79488 ATGCTCTATGAAAAACCAACACATTTATCAACAGAAATTCATCAAGCAATTTGACCGAAAGC 79547
|||
QY 1033 GCTCGGCTTCGAGTTTACACACCGGATATCAGCAGTGGCGGCGGCGGCGGCTCGATG 1092
|||
Db 79548 GCCGATAAACTTGGTCTTAATATATCGAACAATGTTAGCGCGGCGGAGACCGATGCAATG 79607
|||
QY 1093 TTTATCGCCCGAGTCAAGGAGTTCGGAATGTTTTCGTTTCCAAAGCGCTGTCGCCGAGC 1152
|||
Db 79608 ATTTTCGAAAGTTTAAACCAAGTAGGACTTATTTTGTACCAAGCCCAAAAGTATAAGC 79667
|||
QY 1153 CAGGTTCCGGAAGATGGAACCGATTTTCGATGACCTTTCCGCAAAAGGAACTGAGGTGTCCT 1211
|||
Db 79668 CATGCACCAAGAAATGGACCGATTACGACAAAGCTCCAAAAGGATCGAAGTCGTACT 79726
|||
```

RESULT 10

AAQ26728

ID AAQ26728 standard; DNA; 1860 BP.

XX

AC AAQ26728;

XX

DT 28-JAN-1993 (first entry)

XX

DE Heat resistant carbamylase gene.

XX

KW Temp; stable; N-carbamyl-L-Lamino acid; ss.

XX

OS Escherichia coli.

XX Key Location/Qualifiers
FH CDS 320..1550
FT /*tag= a
FT XX
PN JP04183391-A.
XX
XX
PD 30-JUN-1992.
XX
XX 15-NOV-1990; 90JP-00307221.
XX
XX 15-NOV-1990; 90JP-00307221.
XX
XX (NIPS) NIPPON SODA CO.
XX
XX WPI; 1992-265588/32.
DR P-PSDB; AAR25693.
XX
PT Heat resistant carbamylase - is used to produce L-aminoacid in high
PT purity by reacting enzyme protein with N-carbamyl-L-aminoacid.
XX
XX Disclosure; Fig 2; 8pp; Japanese.
XX
XX The DNA encodes a heat resistant carbamylase (optimum temp. ca. 60
XX degrees C, optimum pH ca. 8.5) which can be stably produced by bacteria.
XX The enzyme allows efficient prodn. of a high concn. of N- carbamyl-L-
XX amino acids which may be hydrolysed to the relevant L- amino acids in the
XX presence of the heat resistant carbamylase
XX
SQ Sequence 1860 BP; 438 A; 456 C; 580 G; 386 T; 0 U; 0 Other;

Query Match 14.7%; Score 182.2; DB 2; Length 1860;
Best Local Similarity 48.5%; Pred. No. 7.1e-34;
Matches 564; Conservative 0; Mismatches 593; Indels 6; Gaps 2;

QY 72 AGGCCCCGGTGTACCGGCTGACCTACACTCCAGAGCATGCCGCGCGGGAACGCT 131
DB 382 AAGCGCGCGCTGACGCGCTCTGTTCACTGCTGAAGCGCGGCGCAAGATCTCGT 441
QY 132 CATTCGGCTATGAAGCGCGCGCTTGAGCGTTTCGTGAAGACGCACTCGGAAACATCAT 191
DB 442 CGCTTCTACATGCGCGAAGCGGCTTTTCGTATATGAAGACGCGGCTGCACTTGAT 501
QY 192 CGGCGACGTGAAGCGCACTGATCCGGAGCTTCTCGATCGCGGTGCTACACTTGA 251
DB 502 CGGACGGAAGAGGAGCAATCCGGATGCCAGTCCGCTGCTTGTGATCTCATCTCGA 561
QY 252 TTCTGTCCGAAACGCGCGGATGTTGATGGCACTGCGAGCGTGTGTGCGCCCTTGAGGC 311
DB 562 TTCTGTTTACACGCGCGCTGCTTTGATGACCGCTCGGGTGTGGCGCGGTGGAAGT 621
QY 312 TCCCGGGTGTGCTGGAGAACGGCTACGTGAATCGGCATCCATTTGAGTTTCATCGCAT 371
DB 622 CGTTGACAGCATGAACGAGCAGCGTGTGTGACGCAACACCAATTAAGTAGTGGCGTT 681
QY 372 CGTGAGAGGAAGGGCGCGCTTACAGTGGCATGTGGCGCGCGCGCGCATTTGCAGG 431
DB 682 CACTGACGAAGAGGAGCGCGCTTTCGTTTCGGCATGATCGGACGCGCGCATGGCCGG 741
QY 432 GTTGTGCGCGACGAGCACTGCACTCTTTGTTGATGAGGATGAGTGTCCGTTAGGCA 491
DB 742 AACACTGCC--GCGGAAGCGCTGAGTGCCGCAACGCGGAAGGATTTCCCTCGCTGA 798
QY 492 GCGCGCTACTGCTTCGCTTGAAGCGCGCGCAATGTCAGAGGTGAGCGCGCTTCGCGGC 551
DB 799 AGCGATGAACAGGCGGGCTTGAACCGGACCGCTTGCAGGCGCGCGCAAAACCAAG 858
QY 552 GCACTGCGTCTTTTATCGAATACATTAACAAGACCGATCTTCGACGAGCA 611
DB 859 AACGCTGAAGCGCTATGTCGAATATGATATCAACAAGACCGGTGCTTGAGGACAGG 918
QY 612 AATAGAGATCGGAGTGTGAACCTCCATCTGTCGCGTTTCGCGATTTGCGGTTGCCGTCAA 671

DB 919 ACTTCCAGTTGGCATCGTCACTGGCATCGCGGCTGTGATTTGGGTGAAATTTACCATCGA 978
QY 672 AGGCAGAAGCACACCGCGGCACAAACCCCATGCACCTGCGCCAGGATGCGCTGCTACC 731
DB 979 AGSAAAGCGGAACATGCGCGCGCACCGCATGTCTATTGCGCGCGCACCGATGCGCGC 1038
QY 732 CGCGCTCTCATGTGTAGGGAGGTCAACCGGTTGCTCAACGAGATCGCCGATGGCACAGT 791
DB 1039 GCGCGCCAGATCATATGATGATCGAAGAGCAAGACGAA---CAGGCAACACGCT 1095
QY 792 GGCTACCGTTGGCCACCTCACAGTGGCCCCCGGTGGAGCAACACAGGTCCCGGGGAGGT 851
DB 1096 CGGTACTGTAGACAGTTGCATATATATCCGCGCGGTATTAAATGTCAATTCGGAACGCGT 1155
QY 852 GCACTTCACTGGACCTGCGTTCTCCGATGAGGAGTCCGCTCCGCGTGTGATCGACCG 911
DB 1156 CGAATTTGTGCTCATTTGCGCGACTTGAAGGCTGAGGTGCGCATCAAGTATGGAAGC 1215
QY 912 CATCTCGGTATGCTCGCGAGGTCCGCTCCAGCGCGGTGTGGCTGCCGATGTGGATGA 971
DB 1216 CATAGCCGTGCGGACAGACGATCGCAAGAGCGGNAACGTCCGCGTCAACCGGCGG 1275
QY 972 ATTTTTCATCTACGCCGCGTGCAGTGGTCTTACCATGTGAGACGCGCTTCGCGAAGC 1031
DB 1276 GCTCAAGAAATCCGCGGTGCTATGTTCCGATGAGGTGAAGCGCGCGGAGCGCGC 1335
QY 1032 GGCCTCGGCTTGCAGTTCAACACCGGATATCAGCAGTGGGCGGCGGCGCAGCTCGAT 1091
DB 1336 GTGCCAAAAGCTTGGCTACCGCTGCTTTGCTGCCGAGCGCGCAGCCCATGACAGCT 1395
QY 1092 GTTCATGCCCCAGGTACGGACGTCCGAATGTTTTCGTTCCAAAGCGTGTGCGCGGAG 1151
DB 1396 ACAGTTGCTCCGATTTGCCCAATCGGATGATTTTGTGCTCCCAAGACGAGTGA 1455
QY 1152 CCAGCTTCCGAAAGATGACCGGATTTGATGATGATCTTCGAAAGAACTGAGGTGTCT 1211
DB 1456 TCATAGTCCGCGGAATGGAGTACTAAAGAAGACTGCGCGCTGGAGCAGAGTGTCTTA 1515
QY 1212 CCGGTAATGAGGCACTTGACC 1234
DB 1516 TCATACAGTGTGGCAACTGGCCC 1538

RESULT 11
AAT39170
ID AAT39170 standard; DNA; 1230 BP.
XX
AC AAT39170;
XX AC
DT 16-OCT-2003 (revised)
DT 04-MAR-1997 (first entry)
XX
XX Bacillus stearothermophilus amino acid amidohydrolase DNA.
XX Amino acid amidohydrolase; carbamylase; L-methionine; stereospecific;
KW thermostable; N-carbamyl; cam gene; ds.
XX
XX Geobacillus stearothermophilus.
XX
XX Key Location/Qualifiers
FH CDS 1..1230
FT /*tag= a
FT /product= "amino acid amidohydrolase"
FT /trans= "except= pos:1..3, aa:Met
FT /note= "TTG initiation codon"
XX
XX FR2728905-A1.
XX
XX 05-JUL-1996.
XX
XX 29-DEC-1994; 94FR-00015838.
XX
XX 29-DEC-1994; 94FR-00015838.

XX (RHON) RHONE POULENC NUTRITION ANIMALE.
XX Dion M, Batisse N, Weigel P, Lecocq FM, Hallett JN, Sakanyan V;
XX WPI; 1996-335881/34.
XX P-PSDB; AAW03544.
XX New stereospecific, heat-stable amino acid amidohydrolase - from
PT B.steartotherophilus and related DNA, esp. for prodn. of L-Met.
XX Claim 4; Page 9-11; 17pp; French.
XX Pet1 fragments of Bacillus steartotherophilus NCIB 8224 genomic DNA were
CC inserted into pBR322 and used for transforming E.coli hosts. Selection
CC was on medium which contained tetracycline but lacked arginine. All
CC positive transformants carried plasmids with a 4.7 kb insert. The present
CC sequence is a fragment of the insert. Amino acid amidohydrolase encoded
CC by the fragment is stereospecific and heat-stable (maximum activity at 55
CC -60 deg.C). The enzyme hydrolyses N-carbamoyl amino acid derivs. to L-
CC amino acids and is particularly useful for production of L-methionine.
CC (Updated on 16-Oct-2003 to standardise OS field)
XX
XX Sequence 1230 BP; 268 A; 312 C; 413 G; 237 T; 0 U; 0 Other;
Query Match 14.4%; Score 179; DB 2; Length 1230;
Best Local Similarity 48.3%; Pred. No. 3.8e-33;
Matches 562; Conservative 0; Mismatches 595; Indels 6; Gaps 2;
QY 72 AGCCCCGGTGTACCGCGGTGACCTACACTCAGAGATGCCCGCGCGGGGAACGCT 131
DB 63 AAGCGCGCGGTACCGCGCTCTCGTTCACTGCTGAAGAGCGCGGGCCAAAGATCTCGT 122
QY 132 CATTCGGCTATGAAGCGCGCCTTGAGCGTTCGTGAAGACGACCTCGGAACATCAT 191
DB 123 CGCTTCTACATCGCGAAGCGGGCTTTTCGTATATGAAGACGCGGCTGCGCACTTGAT 182
QY 192 CGCGCGACGTGAAGGACCTGATCCGGAGCTTCTCGCATCGCGCTCGGTTTCACACTTCGA 251
DB 183 CGSACGGAAGAGGACCAATCCGGATGCCAGTCGCTCTTGTGGATCTCATCTCGA 242
QY 252 TTCTGTCCGAACCGCGGATGTTTGAATGCACTGACGGGTGTTGCGCCCTTGAGCG 311
DB 243 TTCTGTTCACACCGCGGCTGCTTTGATGACCGCTCGGGGTGTTGCGCGCGGTGAAGT 302
QY 312 TGCCCGGTTGATCTCGAAGCGCTACGTGATCGGCATCCATTTGATTCATCCGAT 371
DB 303 CGTTACAGCATGAACGACGACCGTGTGTGACGACACCCCAATTGAAGTAGTGCGGT 362
QY 372 CGTGAGGAGGAGGCGCGCTTCAGCATGTGCATGTTGGCGCGCGGCCCATTTGCAGG 431
DB 363 CACTGACGAAGAGGAGCGGCTTTCGTTTCGATGATCGGACGCGGCCATGCGCGG 422
QY 432 GTTGTGCGCGACAGGGAACGTGACTCTTTGTTGATGAGGATGGAGTGTCCGTTAGGCA 491
DB 423 AACACTGCCC---GCCGGAAGCGCTCAGTCCCGACGCGGAAGGATTTCCCTCGCTGA 479
QY 492 GCGGCTACTGCTTCGCTTGAAGCGCGGCAACTGTCAGGCTGACGCGCTCCGCGCG 551
DB 480 AGCGATGAACAGCGCGGGCTTGACCCGACCGCTTCGCCGAGCGCGCAAAACAGG 539
QY 552 GGACCTCGCTGCTTTTATCGAATACACATTTGAACAGGACCGCATCTCGAGCAGGAGCA 611
DB 540 AACGGTGAAGCCTATGTGCAATTTGATATCGAACAAGGACGCGGTCTGGAGGAGCTGG 599
QY 612 AATAGAGATCGGAGTTGTAACCTTCCATCGTTGCGGTTTCGCGCATTCGCGGTTGCCGTCAA 671
DB 600 TCTTCCAGTTGGCATCGTCACTGGCATCGCGGCTCTGATTTGGGTGAAATTTACCATCGC 659
QY 672 AGGCAGAGCGGACCGCGGCAACCCCGCATGACCTGCGCAGCATGCGCTGTGTACC 731
DB 660 CGGCGCGCGGACATCGCGCGGCCACGCGGATGTCAATTGCGCGCGCACCCGATGCGCG 719

QY 732 CCCCCCTCTCATGGTGGAGGAGGTCAACCGGTTCTGTCAACAGATGCGCGATGGCAGCT 791
DB 720 GGCGCGCCAGATCATATAGTATGATCGAAGAGGAGCAAGACGAA---CAGSGACAACGCT 776
QY 792 GGCTACCGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACAGGTCCTCCGGGGAGCT 851
DB 777 CGGTACTGTAGGACAGATTGATATATCCGGCGGTATTATATGTCATTTCCGGAACGGGT 836
QY 852 GGACTTCACACTGGACCTGCGTTCTCCGATGAGAGTGCCTCCGCGTCTGATCGACCG 911
DB 837 CGAATTTGCTCGATTTCGCGCACTTGAAGCTGAGGTGCGGATCAAGTATGGAAGC 896
QY 912 CATTCGCTCATGGTTCGCGAGGTGCGCTCCAGGCCGCTGTGGTGCAGATGTGATGA 971
DB 897 CATAGCGGTGCGGCGAGAGCATGCCAAGAGGAGGAAACGTTTCGCTTCAGCACCGAGG 956
QY 972 ATTTTTCATCTCAGCCCGGTGCGAGTGGCTCCTACCATGTGGAGCGCGTTTCGCGAAGC 1031
DB 957 ACTGCAAGAAATGGCGCGGTGTATTGTTCCGAGTGTGAAACAGGCGCGGAAAGAGC 1016
QY 1032 GGCCTCGGCTTTCAGATTTCACACACCGGATATCAGCAGTGGGGGCGGCGACGACTCGAT 1091
DB 1017 GTGCAAGCAGCTCGGGTATCCGCGTTCGCGAGCGGCGACGCCATGACGCGT 1076
QY 1092 GTTCATGCGCCAGGTACGAGCGTCGGAATGTTTTCGTTCCAAAGCCGTGTGCGCGAG 1151
DB 1077 ACAGTTGGCTCGGATTTGCGCGATGATTTTGTTCGCTCCCAAGACGGGGTGAG 1136
QY 1152 CCACGTTCCCGAAGATGAGCAGGATTCGATGACCTTCGCAAGGAACCTGAGGTTGCTCT 1211
DB 1137 TCATAGTCCGCGGGAATGGAGTACTAAAGAAAGACTGCGCGCTTGGAGCAGAGTGCTTTA 1196
QY 1212 CCGGGTAATGAAGGCACTTGACC 1234
DB 1197 TCATACAGTGTGCAACTGGCCC 1219
RESULT 12
ABQ71004/c
ID ABQ71004 standard; DNA; 3880 BP.
AC ABQ71004;
XX
XX 29-AUG-2003 (revised)
XX 29-AUG-2002 (first entry)
XX
XX Listeria monocytogenes 4b contig DNA sequence #946.
XX
XX Antibacterial; Listeria; food contamination; mutational analysis;
XX infection; ds.
XX Listeria monocytogenes ATCC 19115.
XX
XX WO200228891-A2.
XX
XX 11-APR-2002.
XX
XX 04-OCT-2001; 2001WO-FR003061.
XX
XX 04-OCT-2000; 2000FR-00012697.
XX
XX (INSP) INST PASTEUR.
XX (CNRS) CNRS CENT NAT RECH SCI.
XX
XX Kunst F, Glaser P;
XX
XX WPI; 2002-332479/37.
XX
XX New genomic sequences from Listeria species, useful for detection,
PT treatment and prevention of infection, also related polypeptides,
PT antibodies and modulators.
XX
XX Claim 14; SEQ ID NO 3817; 180pp; French.

Db	3078	CCCTAGCCACAATGGTATAAGCCATGCCCCGAAGAATGGACGGATTACGACAAACTCCA	3019
Qy	1191	CAAAGGAAC TGAGTTGTCTCT	1211
Db	3018	AAAAGGAATCGAAGTCGTACT	2998
RESULT 13			
AAQ15312			
ID	AAQ15312 standard; DNA; 7475 BP.		
XX			
AC	AAQ15312;		

XX	27-AUG-2003	(revised)	
DT	25-MAR-2003	(revised)	
DT	17-MAR-1992	(first entry)	
DT			
XX			Plasmid pHP12 encoding enzyme able to transform 5-subst. hydantoin into

DE N-carbamyl amino acid.
XX
XX Hydrolyase; ss.
KW
XX

Fennell Thomas Sp.

ET CDS 1156.3228
/ *tag= a

FT	/label= ORF 2
FT	3232. .5010
FT	CDS

```
FT      /cag= 2
        /label= ORF 3
```

```
FT /*tag= c
FT /label= OFF 4
```

XX PN JP03251176-A.

08-NOV-1991, PD

PF 29-MAY-1990; 90JP-00137120.
XX

PR 10-JAN-1990; 90JF-00002/86.
XX

XX
XX
XX
XXXX 1001 JANUARY 1974

DR P-PSDB; AAR15482, AAR15483, AAR15484.
XX

PT Enzyme for transforming 5-subst. hydantoin - in which enzymatic protein PT has specified aminoacid sequence.

PS Disclosure; Fig 3; 11pp; Japanese.

CC The sequence comprises three open reading frames: 2, 3, and 4, which
CC together encode an enzyme which has the ability to transform 5-substd.

CC ORFs was obtd. from *Pseudomonas* NS 671 and ligated into pUC18 to form the

CC 27-AUG-2003 to correct OS field.)
vv

SQ Sequence 7475 BP; 2424 A; 1237 C; 1698 G; 2116 T; 0 U; 0 Other;

Best Local Similarity 48.0%; Pred. No. 6.4e-28;

Qv 142 ATGAAAGCGGGCGCTTGAGCGTTCTGTGAAGACGCATCTCGAAACATCATCGGCGGCGT 201

Db 5175 ATGAGAGAGCGCGCCTTACTGTTACTCATGATCATTTTGGGAATTTAAATTGACGGAAA 5234

QY 202 GAAGGCACTGATCCGGAGCTTCC¹TGCGATCGCGGTCGGTT²CACACTTCGATTCTGTCCGA 261

Db 5235 GAGGGAGAAACCCCAAGCCTTCCTTCGTAATGATCGGTTCTCATATTGATTCGTTCGA 5294
Qy 262 AACGGCGGATGTTTGTATGCACTGCGAGCGTGGTGTGCGCCCTTGAGGCTGCCGGTG 321
Db 5295 AATGGCGGTAAGTTTGTATGCGGTTATTGGAGTACTGCGAGGATGAAATTGTCACGCC 5354
Qy 322 ATGCTGAGAACCGCTACGTGAATCGGCATCCATTTGAGTTTCATCGCATCGTGGAGGAG 381
Db 5355 ATATCCGAGGCAAAATGTCGTTTCATGAACATTCATTAAGTGTAGCTTTTGTGAAGAG 5414
Qy 382 GAAGGGCGCGCTTCACAGTGCATGTTGGCGCGCGGCCCATTCGAGGGTTGTTGCC 441
Db 5415 GAAGGTTCAAGGTTTAAATGACGATTATTCGGAAGTAGAGGCATGTTGGAAGGTA 5474
Qy 442 GACAGGAACTGACACTTTTGGTTGATGAGATGGA---GTGTCCGTTAGGCAGGCGGCT 498
Db 5475 CCGAGGATTTACAAAAGTTGACGATACAAATGTAAGTATGATGATGAGCGCTAA 5534
Qy 499 ACTGCTTCGGTTGAAGCGCGGAACTGACGGCTGCAGCGCTCCGCGCGGACCTG 558
Db 5535 TTTGGCTTTGGATAGATCCTGATTTTACCACCAATCTATACGGGAGATTGTTGACATT 5594
Qy 559 CGTCTTTTATCAACTACATGTAACAGGACCGATCTCGACGAGGAGCAATAGAG 618
Db 5595 AACATTTATTTGAGATGCAATTTGAGCAAGGCGCATATCTAGAAAAGTAATATTACCCA 5654
Qy 619 ATCGAGTTGTAACTCCATCGTTGCGTTGCGGCAATGCGGTTGCGTCAAGGCAGA 678
Db 5655 ATAGGATTTGATGCGGATGCGGCTCTCTTGGTTCAAGTAAAGCTAGTTGGAGAA 5714
Qy 679 AGCGACACCGCGGCAACCCCATCTGCGCCAGGATCGCTGTTGTTACCGCGCT 738
Db 5715 GCTGGCCATCGCGGACTGTCTATGAGTCTACGCAAGACCCCTAGTTCGGGCTGCG 5774
Qy 739 CTCATGTTGAGGAGT---CAACCGTTGCTCAACGAGATCCCGATGACAGTGGCT 795
Db 5775 GAAGTCATTAAGAAGTAGAAACTCTTTGATGAACGATCCAAATCTCCACAGTCGGC 5834
Qy 796 ACCTTTGGCCACCTCACAGTGGGCCCCGCTGGAGCAACCCAGTCCCGGGAGTGGAC 855
Db 5835 ACTGTTGTTAGATAGCGGCTTTCTGAGGAGTAATATTAATCTGATCAGTGGAG 5894
Qy 856 TTCACACTGACCTGCTTCCGCAATGAGGATCGCTCGGCTGCTGATCGACCGCATC 915
Db 5895 TTTACCTTTGATATCAGGATATTGAATCGAGAGCAAAACAAATTTATGAGAAGATA 5954
Qy 916 TCGGTCATGTCGGCAGTGCCTCCAGCGCGTGTGCTGCCATGTTGATGATTT 975
Db 5955 GAAGAAAGATCAAACTTTGTAAGTAACACGAGGATTTGGAATPACAGATTGAAAAAAC 6014
Qy 976 TTTCAATCTCAGCCGCTGAGCTGGCTCTTACCATGTTGAGCGCGCTTCGGAAGCGCC 1035
Db 6015 ATGGCTGAGTTCTCTGTAATGCTCTGAAATCTTAATCAATCTCTTAAACATCATGC 6074
Qy 1036 TCGGCTTCAGTTTCACACCGGGATATCAGCAGTGGGCGGCGCCACGACTCGATGTTTC 1095
Db 6075 AAGGAGCTGAGATAGATGACCGATTATCGTTAGTGTGTCAGACACGATGCGATTT 6134
Qy 1096 ATGCCCCAGTTCAGACGTCGGAATGTTTCTGTTCCAGCGCTGCGCGGAGCCAC 1155
Db 6135 TTAGCGGAATAACAGAAATAGGAATGTTGTTGCTCCGTTGATGATTAAGCCAC 6194
Qy 1156 GTTCCGGAAGATGAGACGATTTGATGACCTTCGCAAGGAGTCAAGT 1205
Db 6195 TCACCAAAAGTGGCGGAAATAGATGATATTTCTCACTGGAACTAAGGT 6244

RESULT 14

ABK73017

ID ABK73017 standard; DNA; 1218 BP.

XX

AC ABK73017;

XX

DT 13-AUG-2002 (first entry)
XX Bacillus licheniformis genomic sequence tag (GST) #308.
DE Differential gene expression; genomic sequenced tag; GST;
XX altered culture condition; environmental stress;
KW physiological provocation; ds.
KW Bacillus licheniformis.
XX WO2002291113-A2.
PN 11-APR-2002.
XX 05-OCT-2001; 2001WO-US031437.
XX 06-OCT-2000; 2000US-00680598.
PR 27-MAR-2001; 2001US-0279526P.
XX (NOVO) NOVOZYMES BIOTECH INC.
PA (NOVO) NOVOZYMES AS.
XX Berka R, Clausen IG;
PI WPI; 2002-416684/44.
DR Monitoring differential expression of several genes in first Bacillus cell relative to expression of same genes in one or more second Bacillus cells, by using substrate containing Bacillus genomic sequenced tag array.
XX Claim 4; SEQ ID NO 308; 200pp; English.
XX The invention describes a method of monitoring differential expression of genes in a first Bacillus cell relative to expression of the genes in other Bacillus cells, comprising hybridising labelled nucleic acid probes isolated from Bacillus cells to a substrate containing array of Bacillus genomic sequenced tags (GST), examining the array, and determining relative gene expression by an observed hybridisation reporter signal of a spot in the array. The method is useful for measuring the expression of genes in a first Bacillus cell relative to expression of the same genes in one or more second Bacillus cells. The method is useful for monitoring global expression of several genes from a Bacillus cell, discovering new genes, identifying possible functions of unknown open reading frames and monitoring gene copy number variation and stability. Monitoring changes in expression of genes may be used to provide a representation of the way in which Bacillus cells adapt to changes in culture conditions, environmental stress or other physiological provocation. Extensive follow-up characterisation is unnecessary, when one spot on an array equals one gene or one open reading frame, since sequence information is available. This sequence represents a genomic sequence tag (GST) used in the method of the invention. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at fip.wipo.int/pub/published_pct_sequences
SQ Sequence 1218 BP; 305 A; 268 C; 370 G; 275 T; 0 U; 0 Other;
Query Match 12.1%; Score 150.2; DB 6; Length 1218;
Best Local Similarity 47.3%; Pred. No. 3.2e-26;
Matches 492; Conservative 0; Mismatches 543; Indels 6; Gaps 1;
Qy 180 CGGAAACATCATCGGCGGAGTGAAGGACCTCATCGGAGCTTCCTCGCATCGCGTCGG 239
Db 167 CGGAAATCTGTTGGAGAAATAGAGGTACAGAGAACCGTACGAGCGTACGGGATTTAGC 226
Qy 240 TTCACATTCGATTCGTCCGAAACGGCGGATGTTTGTATGGCACTCGAGGCGTGTGTG 299
Db 227 TTCACATTTGATACGGTTCATCAACGGAGGGAATTCGACGAGCGTACGGGATTTAGC 286
Qy 300 CGCCCTTGAAGGCTGCCGGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTGA 359
Db 287 GAGCCTTTTGTCTGCCAAACACCTTCTTGACACGTCACGCGAGGCCGCAATCAATCGA 346

Db 542 TGACGGGTGTTTTTTCAGAACAGAGATGCAAAAGAGCCAAAGAGACGAGTCGGSAGTCTCCC 601
Qy 485 TTAGGCGAGCGGCTACTGCGCTTGAAGCGCGGAACTGACAGGCTGACAGCCGCT 544
Db 602 TCCAAACAGCGATGATGAGAGCGGTTTGGCAAGGTGTGTTTCACTAGCA-----T 655
Qy 545 CCGCGGGGACCTGCGTGTCTTATCGAACTACACATTTGAACAAGACCGATCTCGAGC 604
Db 656 ATAGGACAGATATCAGCGGCTTGTGTTGAGCTGCATATTGAACAGGGGAAGACGTTGAAA 715
Qy 605 AGSAGCAATAGAGATCGGAGTTGTAACCTCCATCGTTGCGGTTGCGGATTCGGGGTTG 664
Db 716 TGTACGCGCGGATCTCGGATCGTGACAAAGTATTGCGGGGCAGAGACGATATCTCGTCA 775
Qy 665 CCGTCAAAGGAGAGAGACACCGCGGCACAAACCCCATGACCTCGCGCAGGATGCGC 724
Db 776 CGCTCGAAGGAGAAATGCAATCACGACAGGAACCACTCCATGAAATGGCGCAAGACCCGC 835
Qy 725 TGGTACCCGCGCTCTCATGCTGAGGAGGTCAACCGGTTGTCACAGAGATCGCCGATG 784
Db 836 TCGACGCCAGCAGCCGCTATCATGAGCTGCTGCGGTCGGATGAGTGC CGGATG 895
Qy 785 GCACAGTGGCTACCGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACAGGTCCCG 844
Db 896 AGCTCCGCTGACATCGCGAAATAAACGCGAGAGCCCAATGTAGCCAAATGTATACCG 955
Qy 845 GGGAGGTGACCTTCACTGGACCTGCGTTCTCCGCATGAGGAGTCTCGCTCCGCTGTGA 904
Db 956 GCGCGTCCAGTTTCAATCGATATTGCGCATCAGCATCAGCATGTGCTGGACAGTTTC 1015
Qy 905 TCGACCGCATCTCGGTCATGCTGCGGAGGTGCGCTCCAGGCGGCTGCGGCTGCGGATG 964
Db 1016 ATCAGACATGTTGCTTGTATCAACGGCATTTGCCCTACAAAAGGAATTGCGCTGTGA 1075
Qy 965 TGGATCAATTTTCAATCTCAGCGCGGTGCGCTGCTCTTACCATGTTGGAGCCGTTTC 1024
Db 1076 TTGATGAATATATGCGGATAGAGCCTGTGCCGATGGACGAAAGGCTGAAGGCTGCGGCTT 1135
Qy 1025 GCGAAGCGGCTCGGCTTTGCACTTCAACACCGGATATCAGCAGTGGGGGGGCCAG 1084
Db 1136 TTGNAACAGCATTAGAANAACGGCTTCAGCTGTGAGGAATGTTGAGCGAGCAGGCGATG 1195
Qy 1085 ACTCGATGTTTCATCGCCAGGTCACGGACGTGCGAATGTTTTTCTGTTCCAAAGCCGTGCTG 1144
Db 1196 ACGCGCAATGATCGGAAGCGCTATCTCTGTTGATGCTGTTGTTGCGGAGCCGAGGCG 1255
Qy 1145 GCGGAGCCAGTTCCTCCGAAGATGGAC 1172
Db 1256 GCGTCAGCCCACTCACCGAAGGAATATAC 1283

Search completed: May 4, 2004, 13:34:47
Job time : 624 secs

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 13:17:43 ; Search time 130 Seconds
(without alignments)
5289.109 Million cell updates/sec

Title: US-10-045-063-1
Perfect score: 1239
Sequence: 1 GTGACCTCGAGAAAGCGCA.....TGAAGGCACTTGACCGGTA 1239

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 682709 seqs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents NA.*
1: /cgn2_6/prodata/2/ina/5A COMB.seq.*
2: /cgn2_6/prodata/2/ina/5B COMB.seq.*
3: /cgn2_6/prodata/2/ina/6A COMB.seq.*
4: /cgn2_6/prodata/2/ina/6B COMB.seq.*
5: /cgn2_6/prodata/2/ina/PTUS COMB.seq.*
6: /cgn2_6/prodata/2/ina/backfileseq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1239	100.0	1239	4	US-09-285-055-1
2	214.8	17.3	1488	4	US-09-252-991A-6822
3	214.8	17.3	1698	4	US-09-252-991A-6923
4	210.8	17.0	1380	4	US-09-489-039A-2442
5	145.2	11.7	1450	4	US-09-554-998A-1
6	122.6	9.9	1278	4	US-09-134-000C-2571
7	121.2	9.8	1830121	4	US-09-557-884-1
8	121.2	9.8	1830121	4	US-09-643-990A-1
9	104.8	8.5	924	4	US-09-489-039A-4854
10	97	7.8	1290	4	US-09-543-681A-1026
11	93.4	7.5	480	4	US-09-489-039A-2592
12	93.2	7.5	525	4	US-09-252-991A-6967
13	84	6.8	564	4	US-09-252-991A-6968
14	76.8	6.2	339	4	US-09-489-039A-2583
15	56.6	4.6	405	4	US-09-489-039A-4973
16	52.8	4.3	300	4	US-09-489-039A-2448
17	48.8	3.9	381	4	US-09-252-991A-15810
18	48.8	3.9	537	4	US-09-252-991A-15911
19	48.8	3.9	870	4	US-09-252-991A-15882
20	48.8	3.9	1317	4	US-09-252-991A-15836
21	48.8	3.9	2073	4	US-09-252-991A-15779
22	48.6	3.9	1104	4	US-09-252-991A-13171
23	48.6	3.9	1725	4	US-09-252-991A-13170
24	48.2	3.9	1425	4	US-09-252-991A-217
25	48.2	3.9	3114	4	US-09-252-991A-225
26	48.2	3.9	47981	4	US-09-679-279-1
27	47.4	3.8	13613	3	US-09-105-537-3

ALIGNMENTS

RESULT 1

US-09-285-055-1
; Sequence 1, Application US/09285055
; Patent No. 6352848
; GENERAL INFORMATION:
; APPLICANT: ALTENBUCHNER, JOSEF
; APPLICANT: MATTES, RALF
; APPLICANT: PIETZSCH, MARKUS
; APPLICANT: SYLDATK, CHRISTOPH
; APPLICANT: WIESE, ANJA
; APPLICANT: WILMS, BURKARD
; TITLE OF INVENTION: RECOMBINANT L-N-CARBOXYLASE FROM ARTHROBACTER
; TITLE OF INVENTION: THEREWITH
; FILE REFERENCE: RECOMBINANT L-N-CARBOXYLASE
; CURRENT APPLICATION NUMBER: US/09/285,055
; EARLIER FILING DATE: 1999-04-02
; EARLIER APPLICATION NUMBER: DE 198 14 813.5
; EARLIER FILING DATE: 1998-04-02
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Arthrobacter aureus
US-09-285-055-1

Query Match	100.0%	Score 1239;	DB 4;	Length 1239;
Best Local Similarity	100.0%	Pred. No. 5,1e-309;		
Matches 1239;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
QY	1	GTGACCTCGAGAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCGG	60	
DB	1	GTGACCTCGAGAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCGG	60	
QY	61	TTCTCGGAGAGAGCGGCGGCTGTACCGGCTGACCTACCTCCAGAGCATCCGGCGG	120	
DB	61	TTCTCGGAGAGAGCGGCGGCTGTACCGGCTGACCTACCTCCAGAGCATCCGGCGG	120	
QY	121	CGGAAACGCTCATTCGGGCTATGAAAGCGCGGCTTCCGAGCGCTTCCGAGACGACATC	180	
DB	121	CGGAAACGCTCATTCGGGCTATGAAAGCGCGGCTTCCGAGCGCTTCCGAGACGACATC	180	
QY	181	GGAACATCATCCGGCGAGCTGAAAGCACTGATCCGGAGCTTCTCCGATCCGGTCGGT	240	
DB	181	GGAACATCATCCGGCGAGCTGAAAGCACTGATCCGGAGCTTCTCCGATCCGGTCGGT	240	
QY	241	TCACATTCGATCTGTCCGAAACCGCGGGATGTTTGTGGCACTCCAGCGTGTGTGC	300	
DB	241	TCACATTCGATCTGTCCGAAACCGCGGGATGTTTGTGGCACTCCAGCGTGTGTGC	300	

```
QY 301 GCCCTTAGCGTCCCGGTGATGCTGGAGACGGCTAGTGAATCGGCATCCATTGAG 360
Db 301 GCCCTTAGCGTCCCGGTGATGCTGGAGACGGCTAGTGAATCGGCATCCATTGAG 360
QY 361 TTTCATCGCGATCGTGAGAGAGAAAGGGGCCGCTTCAGCAGTGGCATGTGTGGCGCGCGG 420
Db 361 TTTCATCGCGATCGTGAGAGAGAAAGGGGCCGCTTCAGCAGTGGCATGTGTGGCGCGCGG 420
QY 421 GCCATTGCAAGGTGGTCCCGACAGGGAACCTGGACTCTTTGGTTGATGAGATGAGGTG 480
Db 421 GCCATTGCAAGGTGGTCCCGACAGGGAACCTGGACTCTTTGGTTGATGAGATGAGGTG 480
QY 481 TCGGTTAGCAGCGGCTACTGCTTCGCTTGAAGCGGGCGAACTGCAGGCTCAGCC 540
Db 481 TCGGTTAGCAGCGGCTACTGCTTCGCTTGAAGCGGGCGAACTGCAGGCTCAGCC 540
QY 541 CGCTCCGCGCGGACCTCGGTGCTTTTATCGAACTACATATTGAACAAGACCGATCCTC 600
Db 541 CGCTCCGCGCGGACCTCGGTGCTTTTATCGAACTACATATTGAACAAGACCGATCCTC 600
QY 601 GAGCAGGACAAATAGAGATCGGAGTTGTAACTCTCATCGTTGGCGTTTCGCGCATTTGGG 660
Db 601 GAGCAGGACAAATAGAGATCGGAGTTGTAACTCTCATCGTTGGCGTTTCGCGCATTTGGG 660
QY 661 GTTGGCGTCAAGGAGAGACGACCGCGGACAAACCCCATGCACCTGCGCCAGGAT 720
Db 661 GTTGGCGTCAAGGAGAGACGACCGCGGACAAACCCCATGCACCTGCGCCAGGAT 720
QY 721 GCGCTGGTACCCGCGCTCTCATGTGAGGAGGTCAACCGGTTTGTCAACGAGATCGCC 780
Db 721 GCGCTGGTACCCGCGCTCTCATGTGAGGAGGTCAACCGGTTTGTCAACGAGATCGCC 780
QY 781 GATGACAGTGGTACCTTGCGCCACTCA CAGTGGCCCCCGGTGGAGGCAAC CAGGTG 840
Db 781 GATGACAGTGGTACCTTGCGCCACTCA CAGTGGCCCCCGGTGGAGGCAAC CAGGTG 840
QY 841 CCGGGGAGGTGGACTTCACTGGACCTGCGTCTCCGATGAGGAGTCCGTCGCGTG 900
Db 841 CCGGGGAGGTGGACTTCACTGGACCTGCGTCTCCGATGAGGAGTCCGTCGCGTG 900
QY 901 CTGATCGACCGCATCTCGGTGATGTCGGCGAGGTGCGCTCCAGCGCGGTGTGGCTGCC 960
Db 901 CTGATCGACCGCATCTCGGTGATGTCGGCGAGGTGCGCTCCAGCGCGGTGTGGCTGCC 960
QY 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGCACTGGCTCTTACCATGTGGAACGCC 1020
Db 961 GATGTGGATGAATTTTCAATCTCAGCCCGGTGCACTGGCTCTTACCATGTGGAACGCC 1020
QY 1021 GTTCGGAGCGGCTCCGCTTGCAGTTTCA CACACCGGGATATCAGCAGTGGGGCGGGC 1080
Db 1021 GTTCGGAGCGGCTCCGCTTGCAGTTTCA CACACCGGGATATCAGCAGTGGGGCGGGC 1080
QY 1081 CACGACTCGATGTTTATCGCCCGAGTCA CCGAGCTCGGAATGGTTTTGTTCCAAAGCGT 1140
Db 1081 CACGACTCGATGTTTATCGCCCGAGTCA CCGAGCTCGGAATGGTTTTGTTCCAAAGCGT 1140
QY 1141 GCTGCGGAGCCACGTTCCGGAAGATGGA CCGATTTGATGACCTTCGCAAAAGAACT 1200
Db 1141 GCTGCGGAGCCACGTTCCGGAAGATGGA CCGATTTGATGACCTTCGCAAAAGAACT 1200
QY 1201 GAGGTTGCTCCGGTAAATGAAGCACTTGACCGGTAA 1239
Db 1201 GAGGTTGCTCCGGTAAATGAAGCACTTGACCGGTAA 1239
```

RESULT 2

```
US-09-252-991A-6822
; Sequence 6822, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
```

```
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
```

```
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6822
; LENGTH: 1488
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6822
```

```
Query Match 17.3%; Score 214.8; DB 4; Length 1488;
Best Local Similarity 50.8%; Pred. No. 7.1e-46;
Matches 547; Conservative 0; Mismatches 517; Indels 12; Gaps 1;

QY 146 AAGCGGCGCCTTGAGCGTTCTGTGAAGACGCACTCGGAAACATCATCGCGCAGCGTGAAG 205
Db 380 AAGCGGCGGCTGCAACGGTCAGCGTCAGGTCGGGTGCGCAACATCTTCGCCCGTCTGCCG 439
QY 206 GCACCTGATCCGAGAGTTCCTGCGATCGCGTCTCGGTTCACACTTCGATTCTGTCCGAAACG 265
Db 440 GCGGTAAACCCGACTTGGCCCCGGTGATGACCGGTAGCCACATCGACACCCAGCCACCG 499
QY 266 GCGGATGTTGATGGCACTGACAGCGTGTGTGGCCCTTGAGGCTCCCGGGTGATGC 325
Db 500 GTGGCAAGTTCCAGCGGCTGCTTCGGGGTGATGGCGCGCTCGAGGTGATCGGCACCTCA 559
QY 326 TGGGAACGGCTAGTGAATCGGCATCGGCATCCATTGTGATTTCATCGCATCGTGAGGAGGAAG 385
Db 560 ACGACTTCGGGGTGAAACCCAGCGCGCTGAGAGTGGTGTGGAACCAACGAGGAAG 619
QY 386 GGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCCGCGCATTTGCGAGGGTGTGTCGCCGACA 445
Db 620 GCTCGCGCTTCGCGCCCTGCATGATGGCTCGGGGTATTTCGCGGGAAGTTTCACTCTGG 679
QY 446 GGGAACTGGAATCTTTTGGTTGATGAGATGAGATGTCCGTTAGGAGCGCGGTACTGCT 505
Db 680 AGGAGACCTTGCCCAAGCGCGATGCGACGGTGTGAGGTAGGCGAGCGCTGGACGCCA 739
QY 506 TCGGCTTGAACCGGGCGAATCTGACAGGTGTCAGCCCGCTCCGCGCGGACCTGCGTGT 565
Db 740 TCGGCT-----ACGCGGAGCGCGCATTTGTCTCGGACATCCGGTGGCGCT 787
QY 566 TTATCGAACTACATTTGAAACAAGGACCGATCTCTCGAGCAGGAGCAAAATAGAGATCGGAG 625
Db 788 ATTTTCGAGGCGCACATCGAACAGGGCGGATCTCTCGAGGACGAGGAGAAGACCATCGCGC 847
QY 626 TTGTAACTTCATCGTTGGCGTTTCGCGATTTGGGGTGGCGTCAAGGCAAGAGCGACC 685
Db 848 TGGTCTCGGCGCGCTCGGGCAGAAATGGTTTCGACCTGTCTCCGCGCGCTCGAGGGCAC 907
QY 686 ACGCGGCACAAACCCCATGCACTGCGCCAGGATGCGTGTGCTACCCCGCTCTCATGG 745
Db 908 ACGCGGCGCAACCGCGATGCACTTGCCTCAAGGATGCCCTGCTGGTCCCGCGCGGTGG 967
QY 746 TGAGGAGGTCAACCGGTTTCGTCACAGAGATCGCCGATGGCAGATGGCTACCGTTGGCC 805
Db 968 TCGAGGGGTCAATCGCGCGCGCTTCGCGCATAGCCGCGATGCTTTCGCGCAGCGTCCGCT 1027
QY 806 ACCTCAGATGGCGCGCGGTGGAGCAACAGGTTCGCGGGGAGGTGCACTTCACACTGG 865
Db 1028 GCCTGCACGCTATCCCGGTTTCGCGCAACGCTGATACCCGCGCAAGTGAAGATGACCTCG 1087
QY 866 ACCTGCGTTCTCGCATGAGGAGTCCGCTCCGCTGCTGATCGACCGCATCTCGGTCATGG 925
Db 1088 ACTTCCGCCATCTGCAACCGGAGGCGCTTGGACTCGATGATGCCGAAGTTCGCGCAGGTGA 1147
QY 926 TCGGCGAGGTGCGCTCCAGCGCGGTGTGGCTCGCGATGTGGATGAATTTTCAATCTCA 985
```

Db 1148 TCAGCGCTACCTCGGAGAGCATGGCTTGGCAATACGAGCTGGTGCAGACCGCCGATTTCC 1207
Qy 986 GCCCGGTGACGTGGCTCTTACCATGTGGACCGCGTTTCGCGAAGCGGCTTCGGCCTTGC 1045
Db 1208 CGCGCTGTACTTCGACCGAGGATGCGTCGGCGGCTGCGCGAGCGCGCGCTGG 1267
Qy 1046 AGTTACACACCGGATATCAGCAGTGGCGGCGGCGCACGACTCGATGTTTCATCGCCAGG 1105
Db 1268 GCATGCGCAGATGAGACATCGTCAGCGCGCGCGCCACGACGCGATCTTCTCGCGGAA 1327
Qy 1106 TCAGGACGTCGGAATGTTTTCGTTCCAGCGCTGCTGGCCGAGCCAGTTCCCGNAG 1165
Db 1328 TCGGTCCGCGGGAGTATCTTCTGTCCTCCAGAACGCGCATCAGCCACACGAGATCG 1387
Qy 1166 AATGAGCCGATTTTCGATGACTTCGCAAGGAACTGAGGTGTCTCCCGGGTAATG 1221
Db 1388 AGAAGCCAGCCCGACGACCTGGCGCGGCTGCGGCTGCTGTTGGGGCCATG 1443

RESULT 3

US-09-252-991A-6923/c
; Sequence 6923, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6923
; LENGTH: 1698
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6923

Query Match 17.3%; Score 214.8; DB 4; Length 1698;
Best Local Similarity 50.8%; Pred. No. 7.4e-46;
Matches 547; Conservative 0; Mismatches 517; Indels 12; Gaps 1;

Qy 146 AAGCGCCGCTTGAGCGTTTCGTAAGACGCACTCGGAAACATCATCGGCCGACGTGAAG 205
Db 1373 AAGCGCCGCTGCACGGTCAGGTCGATCGGTGCGGCAACATCTTCGCCCGTGTCCCG 1314
Qy 206 GCACATGATCCGAGCTTCCTCGGATCGCGTCCGTTTCACACTTCGATTTCTGTCGGAACG 265
Db 1313 GCGTAACCGGACTTGGCCCGGTGATGACCGGTAGCCAGATCGACACCCAGCCACCG 1254
Qy 266 GCGGATGTTGATGCACTGACGCGTGTGCGCCCTTGAGGCTGCGCGGTGATGC 325
Db 1253 GTGGAAAGTTCGACGCTGCTTCGGGGTGTATGCGCGGCTTCGAGGTGATCCGCCCTCA 1194
Qy 326 TGAGAAACGGTACGTGAATCGGCATCCATTTGAGTTTCATCGCATCGTGGAGAGGAAG 385
Db 1193 ACGACCTCGGGTGGAAACGAGGCCCGCTGAGGTGGTGTGTGACCAACGAGGAAG 1134
Qy 386 GGGCCCGCTTCAGCACTGGCATGTTGGGGCGCGGCCCATTTGCAGGGTTGGTTCGCCGACA 445
Db 1133 GCTCGGCTTCGGCCCTGCATGATGGCTCGGGCGTATTTCGCCGGAAGTTTCAACCTCG 1074
Qy 446 GGGAACTGCACTTTTGGTTGATGAGGATGGAGTTCGCTTAGCGAGCGGCTACTGCTCT 505
Db 1073 AGGAGACCCCTGGCCAAAGCGCATGCGGAGCGGTGTACGCGTAGCGGAGCGGCTGGACGCCA 1014
Qy 506 TCGGCTTGAAGCGGCGCACTGCAAGCTGCAGCCGCTCCGCGGCGGACCTGCGGTCTT 565
Db 1013 TCGGCTT-----ACGCGGGAGCGCGGATTTGTTCTCGGACATCCGCGTGGCGGCT 966

Qy 566 TTATCGAACTACACATTGAACAGGACCGATCTCGAGCAGGAGCAAAATAGAGATCGGAG 625
Db 965 ATTTCGAGCGCACATCGAAACAGGGCCGATCTCTGAGGACGAGGAGAACCATCGGCG 906
Qy 626 TTGTAACCTCCATCGTTGGCGTTTCGGCGATTGGGGTTGCCGTTCAAGGAGGAGCGACC 685
Db 905 TGTGTCGCGCGCTCGGGCAGAAATGTTTCGACCTGTCCCTGCGCGGCTCGAGGCAC 846
Qy 686 AGCCGCGCAAAACCCCATGACCTCGCGCAGGATCGCTGGTATCCCGCGCTCTCATGG 745
Db 845 AGCCGCGCGCAACGCGCATGACCTCGCGCAAGGATGCTCTGTCGCTGCGCGCGGCTGG 786
Qy 746 TGAGGAGGTCAACCGGTTTCGTAACGAGATCGCGATGGCACAGTGGGTACCGTTGGCC 805
Db 785 TGAGGCGGTCAATCGCGCGGCTTCGCGCATCAGCCGATGTTTTCGCGCACGCTCGGCT 726
Qy 806 ACTCACAGTGGCGCGCGTGGAGCAACAGGTCCCGGGGAGGTGGATTCACACTGG 865
Db 725 GCTGACGCTTATCCCGGTTTCGCGCAACGTGATACCCGCGAAGTGAAGATGACCTGG 666
Qy 866 ACCTGCTTCTCGCATGAGGAGTTCGCTCCGGTGTGATCGACCGCATCTCGGTCTATGG 925
Db 665 ACTTCGCGCATCTGCAACCGGAGCGCTCGACTCGATGATCGCCGAAGTCCGCCAGTGA 606
Qy 926 TCGCGAGGTGCGCTCCAGCGCGGTGCTGCGGTCGCGATGATGAATTTTCAATCTCA 985
Db 605 TCGCGCTACCTCGCAGAAAGCATGGCTTCAATACAGAGCTGGTGGCGCACCGCGATTTC 546
Qy 986 GCGCGGTGCGCTGCTCTACCTACCTGCGAGCGGTTTCGCGAAGCGGCTCGGCTTGC 1045
Db 545 CGCGCGTGTACTTCGACAGGAGTTCGTCGCGCGGTGCGGAGCGCGCGCGAGCGCTGG 486
Qy 1046 AGTTACACACCGGATATCAGCAGTGGCGCGGCGCACACTCGATGTTTCATCGCCAGG 1105
Db 485 GCATGCGCGAGATGACATCTGACGCGCGCGCGCACACGCGATCTTCTCGCGAAC 426
Qy 1106 TCAGGAGCTCGGAATGTTTTCGTTCCAAAGCGGTGCTGCGCGGAGCCACGTTCCCGAAG 1165
Db 425 TCGGTCCGCGGGATGATCTTCGTGCTCTGCGAAGCGCATCAGCCACACAGAGATCG 366
Qy 1166 AATGACCGATTTTCGATGACCTTCGCAAGGAACTGAGGTGCTCTCCGGTATG 1221
Db 365 AGAAGCCAGCCCGACGACCTGGCGCGGCTGCGCGGTGCTGTTTCGCGGCGCATG 310

RESULT 4

US-09-489-039A-2442
; Sequence 2442, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2442
; LENGTH: 1380
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2442

Query Match 17.0%; Score 210.8; DB 4; Length 1380;
Best Local Similarity 49.4%; Pred. No. 7.4e-45;
Matches 576; Conservative 0; Mismatches 587; Indels 3; Gaps 1;
Qy 51 GCTCTCCCGGTTCTCGGAGAGGCGCGGTTTACCGCGGTGACCTACACTCCAGACA 110
Db 192 GCTGCGCGCTTTCAGGAAACCGCGACGCCCTGACCGGCTCTATCTGCGCGCGAGCA 251

QY 665 CCCTCAAGGCGAAGGACCAAGCGGCGCAACACCCCATGACCTGCGCCAGATGCGC 724
Db 776 CGCTCGAAGGAGAAATGCAATCAGCGAGGAACACCTCCATGAATGCGCAAGGACCGC 835
QY 725 TGGTACCCCGCTCTCATGCTGAGGAGGTCAACGGTTCTGTCACAGATGCGCGATG 784
Db 836 TCGACGCGAGCAGCGTATCATTCATGAGTGTCTGCTGCGGTGCGATGAGTCCCGATG 895
QY 785 GCACAGTGGCTACGTTGGCCACCTCACAGTGGCCCGCGGTGAGGCAACACAGTCCCGG 844
Db 896 AGCTCGTCTGACATGCGGAATAAATACGCGAGAGCCCAATGTAGCCAATGTATACCGG 955
QY 845 GGGAGTGGACTTCACTAGGACCTGCTGCTCCGATGAGAGTGTCTCCGCGTGTGTA 904
Db 956 GCGCGTCCAGTTTCAATCGATATTCGCCATCAGCATCAGCATGTGCTGGAACAGTTTC 1015
QY 905 TCGACCGCATCTCGTTCATGCTCGGCGAGGTGCGCTCCAGGCGGTGCTGCGCATG 964
Db 1016 ATCAAGACATGTTGCTTGTATCAACGGCAATTTGCCCTACAAAAGGAATTCGCGCTGTGA 1075
QY 965 TGGATGAATTTTCAATCTCAGCCCGTGCAGCTGCGTCTTACCATGCTGGAGCGCGTTTC 1024
Db 1076 TTGATGAATATATGCGGATAGACCTGTGCCGATGAGCAAGGCTGAGGCTGCGCTT 1135
QY 1025 GCGAAGCGGCTCGGCTTGCAGTTCACACACCGGATATCAGAGTGGGCGGCGCACG 1084
Db 1136 TTGAACAGCATTTAGAAAACGGCTTCAGCTGTGAGGAAATGGTGAGCGGAGCAGGCATG 1195
QY 1085 ACTCGATGTTTCATCGCCCGACGTCACGAGCTCGGAATGTTTTCGTTCCAGCGGTGCTG 1144
Db 1196 ACCGCAATATGCGAAGCGCTATCCTGTTGTATGCTGTTGTGCGAGCGGCGG 1255
QY 1145 GCGGAGGACCGCTTCCGGAAGATGGAC 1172
Db 1256 GCGTACGCCACTCACCGAGGATATAC 1283

RESULT 6

US-09-134-000C-2571
; Sequence 2571, Application US/09134000C
; Patent No. 6617156
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; FILE OF INVENTION: ENTEROCOCCUS FAECALIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 032796-032
; CURRENT APPLICATION NUMBER: US/09/134,000C
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/055,778
; PRIOR FILING DATE: 1997-08-15
; NUMBER OF SEQ ID NOS: 6812
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2571
; LENGTH: 1278
; TYPE: DNA
; ORGANISM: Enterococcus faecalis
US-09-134-000C-2571

Query Match 9.9%; Score 122.6; DB 4; Length 1278;

Best Local Similarity 45.9%; Pred. No. 3.3e-22;

Matches 474; Conservative 0; Mismatches 554; Indels 9; Gaps 1;

QY 172 GAGCGACTCGGAACATCATCGCGCAGCGTGAAGGCACTGATCCGGAGCTTCTCGGATC 231
Db 217 GAGCAAGTCGGNAATCTGTTGTGAGTGGCGAGGACCGAATTTCTCAGGAAACGATT 276
QY 232 GCGGTGCGTTCACATTCGATTCGTCCGAAACCGCGGATGTTTGTAGTGCATCGAGCG 291
Db 277 TTAACCTGGGTACATATCAGACAGTAGTTAATGTTGGGACCTTAGATGTCATATGTT 336
QY 292 GTGGTGTGCGCTTCAGGCTGCGCGGTGATGCTCGAGACGCGCTACGTGATCGGCAT 351
Db 337 GTCAATGCTCCATGACAGCTGTTCAATATTTATTAGACAAATATCGGAAACCGCGCT 396

QY 352 CCATTTGAGTTTCATCGGATCGTGGAGGAGGAGGCGCGCTTTCAGAGTGGCATGTTG 411
Db 397 TCTTTAGAAAGTCTTCCATGGCTGAAGAAGGAGTCTGTTTCCACGGTTTCTCG 456
QY 412 GCGCGCGCGGCGCATTCAGAGGTTGGTGGCGGACAGGGAATGGACTCTTGGTGTGATGAG 471
Db 457 GGAAGCAAAAATGTCGTTGTTGAAGCAAAACGCTGAAGAAGTAGTCGATATCTCTGACAAC 516
QY 472 GATGGAGTGTCCGTTAGGCGAGCGGCTACTGCTTCGGCTTGAAGCGGCGCAACTGCAG 531
Db 517 GAAGGTAAAGTTTGTGATGAGATGCAATCGCAAGGCTTTGATTTTCGCAAGAGCAA 576
QY 532 GTGTCAGCCCGCTCCCGCGGCGACCTGCGTCTTTTATCGAACTACACATTTGAACAAGGA 591
Db 577 ACTAGAC-----GTGAGGACATTTGAAGCAATTTGTGCAAAATTCATATTGAACAAGGC 627
QY 592 CCGATCCTCGAGCAGAGCAAAATAGAGATCGGAGTTGTAACTCTCCATCGTGGCGTTGCG 651
Db 628 AATAATTTAGAGAAACGAAAGCTCCAAAGTCCGTTGTTTTCATAGTATTTGAGGCAACGT 687
QY 652 GCATTCGGGTTGCCGCTCAAGGCAAGCGACCGCGGCAACACCCCATGCACTG 711
Db 688 CGGTACACCGTTAACTTTAAAGGCGCAAGCCAATCATGCAAGTACTAGCCCAATGAGTTAC 747
QY 712 CGCCAGGATCGCTGTTACCCCGCTCTCATGCTGAGGAGGTCAACCGGTTTCGTCACAC 771
Db 748 CGTCATGATCGGTGTACGCTTTGCTAAATTTGTGCAAGAGCCATTTGATCGTGCAGAC 807
QY 772 GAGATCGCGATGGCACAGTGGCTACGTTGGCCACTCACAGTGGCGCCCGGTGAGGCG 831
Db 808 GAAATTTGGGAGTCCGTTAGTTTAAATTCGCGCAAGTCAATTCGGAACCCCAATACAGTC 867
QY 832 AACCAGTCCCGGGGAGGTGGACTTTCACACTGGACTGCTTCTCCGATGAGGAGTCG 891
Db 868 AACGTGCTGCGAGAGTGGAAATTTCAATTTGACTGCGCTCATACCGACGCTGCTTTT 927
QY 892 CTCGCGTCTGATCGACCGCATCTCGTCTGCTGCGGAGGTGCGCTCCAGGCGCGT 951
Db 928 TTACGACATTTTACAGTGAATTAGAAGAACGGATGAAGACAATTCCTCAAGAAATTAGGC 987
QY 952 GTGGTGGCGATGTGGATGAAATTTTCAATCTCAGCCCGGTGAGTGGCTCTACATG 1011
Db 988 TTGACGATTCGATGATGCTGCTGATGATGAAGCCCTGTACCGATGAATCAAGCAAT 1047
QY 1012 GTGAGCGCTTCCGAGCGGCTCGGCTTTCAGTTCACACACCGGATATCAGCAGT 1071
Db 1048 GTAGAAGTTATTGAACAAAGCCAAAGAGCAAGTATCAGTATCGTATGATGATAGT 1107
QY 1072 GGGCGGCGCAGACTCGATGTTTCATCGCCAGGTCAAGGACGTCGGAATGTTTTCGTT 1131
Db 1108 GGTGCTGGCATGATTCGCAATTTATGCGCTCATATCCCGACAGCGATGATTTGTG 1167
QY 1132 CCAAGCGGTGCTGGCGGAGCGACGTTCCGGAAGATGACCGATTCGATGACCTTCGCG 1191
Db 1168 CCAAGTATTGTTGGGATAGTATATCTCTGAGAAAGCAACGCGCATTTCCAGATTTGGTT 1227
QY 1192 AAAGGAAGTGGTTGCT 1208
Db 1228 GAAGGTGTCAAGCTTTT 1244

RESULT 7

US-09-557-884-1
; Sequence 1, Application US/09557884
; Patent No. 6506581
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: The Nucleotide sequence of
; the Haemophilus influenzae Rd Genome, Fragments
; Thereof, and Uses Thereof
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:

ADDRESSER: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:
MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/557,884
FILING DATE: 25-Apr-2000

CLASSIFICATION DATA: <unknown>

PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: JUN-5-1995

ATTORNEY/AGENT INFORMATION:
NAME: Michelle S. Marks

REGISTRATION NUMBER: 41,971
REFERENCE/DOCKET NUMBER: PB186P3

TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-309-8504
TELEFAX: 301-309-8439

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear

SEQUENCE DESCRIPTION: SEQ ID NO: 1:

US-09-557-884-1

Query Match 9.8%; Score 121.2; DB 4; Length 1830121;

Best Local Similarity 43.8%; Pred. No. 7.1e-21;

Matches 525; Conservative 0; Mismatches 673; Indels 0; Gaps 0;

QY	28	CGCATGGAAGAGATCGGAGCTCTCCGGTCTCGGAGAGAGCCCGGCTTACC	87
DB	608660	CGTGTCAAATTTAATTGAAACCTGGCTTTTATTCATCAGTACCAACAGCTGACT	608719
QY	88	CGGCTGACCTACACTCCAGCATGCCGCGCGGAAACCTCATTCGCGCTATGAAA	147
DB	608720	CCCTTAGCTTTTACAGAGAGATGAAGGCCAATATGATTTAATGAAATATGAAA	608779
QY	148	GGCGCGCGCTCGAGCGTTCTGTAAGACGCATCGGAAACATCATCGCGCAGCGTGAAGC	207
DB	608780	GAATATGATTTGTCTATTCTGTCGAGATTCATTTGGAATCTTTTATTCTGTAAGGCAGT	608839
QY	208	ACTGATCCGAGCTTCCTGCGATCGGTCGGTTACACTTCGATTCGTCCGAAACGCC	267
DB	608840	AAAGAAGATTTTTCCTGCGATTCGATTCGAGTACATATTTGATCTGTTGTAATGCT	608899
QY	268	GGGATGTTTGATGGCACTCGAGCGTGTGTCGGCTTGAGGCTGCCGCGGTGATGCTG	327
DB	608900	GGTAATTTGATGGTCTTTGGATCTGTGTCGGTGGAAATCTTTTGGCAATTTGT	608959
QY	328	GAGAACCGCTACGTGAATCGGCATCCATTTGATTTTCATCGCATCGTGAGGAGAAAGG	387
DB	608960	GAACAGAAATATTCAGACTCGTTATCTTTAGAAATTTGATAATTTTACTTGTGAAGATCG	609019
QY	388	GCCCGCTTCAGCAGTGGCATTTGGCGCGCGGCGCATTCAGGGTTGTCGCCACAGG	447
DB	609020	AGTAGATTAATTTTGTCTACATTTGGGTAGTAAAGTTATGTTGGCATAGTAATCAAGAA	609079
QY	448	GAACTGGACTCTTTGGTTCATGAGGATGGAGTGTCCGTTAGGACGCGGCTACTCCCTTC	507
DB	609080	AAATTAAGTTTCACTCGTGATAAACAGGAAGGGTTTATCAGAAGCTATGCTGAAGTA	609139
QY	508	GGCTTGAAGCCGGGAACTGAGGCTGACGCCCGCTCGCGCGGAGCCTGCGTCTTTT	567
DB	609140	GGAATGAATTTTAAATTTGGTTAATCAAGCAAAACCGTGTGCAAGGAATTTAAATGTTTT	609199

QY	568	ATCGAACTACACATTTGAACAAGGACCGATCCTCGAGCAGGAGCAATAAGATCGAGTT	627
DB	609200	TTTGAATTTATATAGAACAGGCCCTCGTTTGAAGATGAGGAAACATAGTGT	609259
QY	628	GTAACCTCCATCGTTGGCGTTTCGCGCATTTGCGGTTGCGGTCAAAGGAGAGCGACAC	687
DB	609260	GTGACAGGTAATTGCTGCCCGCATTCGTGCAATTTGTTAAATTTAAAGCAAGCGGATCAT	609319
QY	688	GCCGCAACACCCCATGACCTGCGGCAGAGTGGTGGTACCGCGCTCTCATGTG	747
DB	609320	TCAGAGCAACAGCAATGCAATTTATCGTCATGATGCAATTTGTTAGGAGGATCTGAGTTATCA	609379
QY	748	AGGAGGTCAACCGGTTCTGTCACAGATGCCCGATGCGCATGCGTACCGTTCCGCCAC	807
DB	609380	CTTGCTATTGAGCGAGCTGCTATTCAAGCTGGACATTTCTACAGTAGCTACAGTAGTAAT	609439
QY	808	CTCAGTGGCCCGCGTGGAGCAACAGAGTCCCGGGGAGGTGGACTTCCACTCGAC	867
DB	609440	ATTACAGCTAAACCCAGGAGTAATGAAATGTTGTCAGGATATTGCGAATTTATTAGTAGAT	609499
QY	868	CTGGCTTCGCGATGAGGAGTGGTCCGCGTGTGCTGATCGACCGCATCTCGTCAATGTC	927
DB	609500	ATTCAGGTACACATGTACAAGCTAGAGATTTCTGTATTGTAATTTATCAAGAGGAAT	609559
QY	928	GGCAGGTGCGCTCCAGGCCGCTGTGCTGCCGATGTGGATGAATTTTCAATCTCAGC	987
DB	609560	AGTAAAGTTTCAGAAAAGAGAGATTTAATTGATTGATTACACTTATTTCAAGATAAT	609619
QY	988	CCGCTGACGCTGGCTCTTACCATGGTGAACCCGCTTCGCGAAGCGGCTCGGCTTGCAG	1047
DB	609620	CCAAATATTACCTGAAATATGTAACCAATAGTGAACAGCAGACATTTCCCTTGGT	609679
QY	1048	TTACACACCCGGGATATCAGCAGTGGGCGCGCCACGACTCGATGTTTCATCGCCAGTC	1107
DB	609680	TATTTCTACGAAATTAATGCCAAGTGGTGCAGGACATGATGCAATGCAATATGGCAACT	609739
QY	1108	ACGAGCGTGAATGTTTTCGTTCCAGCGCTGCTGCGCGGAGCCAGCTTCCCGAAGAA	1167
DB	609740	TGCCAACAGATGATGATTTTATTCATCTCATTTGGGAATTAGTCATAACCTCTTGA	609799
QY	1168	TGGACCGATTCGATGACCTTCGCAAGGAACTCAGGTTGTCCTCCGGTAAATGAAG	1225
DB	609800	TTTACTGATGGAAGATATAGAGCGAGGAATTAAGTTTTTACAAAAAGTTATATTGG	609857

RESULT 8

US-09-643-990A-1
Sequence 1, Application US/09643990A
Patent No. 6528289

GENERAL INFORMATION:

APPLICANT: Robert D. Fleischmann
Mark D. Adams
Owen White
Hamilton O. Smith
J. Craig Venter

TITLE OF INVENTION: The Nucleotide sequence of
the Haemophilus influenzae Rd Genome, Fragments
Thereof, and Uses thereof

NUMBER OF SEQUENCES: 1

CORRESPONDENCE ADDRESS:

ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville,
STATE: MD
COUNTRY: USA
ZIP: 20850

COMPUTER READABLE FORM:

MEDIUM TYPE: 3 1/2 inch diskette
COMPUTER: Dell Pentium
OPERATING SYSTEM: MS DOS v6.22
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/643,990A
FILING DATE: 23-Aug-2000
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: 1995-06-07
APPLICATION NUMBER: 08/426,787
FILING DATE: 1995-04-21
ATTORNEY/AGENT INFORMATION:
NAME: Kenley K. Hoover
REGISTRATION NUMBER: 40,302
REFERENCE/DOCKET NUMBER: PB186P1C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 301-610-5790
TELEFAX: 310-309-8439
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1830121 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
SEQUENCE DESCRIPTION: SEQ ID NO: 1:
US-09-643-990A-1

Query Match 9.8%; Score 121.2; DB 4; Length 1830121;
Best Local Similarity 43.8%; Pred. No. 7.1e-21;
Matches 525; Conservative 0; Mismatches 673; Indels 0; Gaps 0;

QY 28 CGCATTTGAGAAAGAGATCGGGAGCTCCCGGTTCTCGGCAGAAAGCCCGGTGTACC 87
DB 608660 CGTGTTCAAAATTTAATGAAAACCTGGCTTTTATTTCATCATGACCAACGAGCTGACT 608719

QY 88 CGGCTGACCTTACACTCCAGAGATGCGCGCGCGGAAACGCTCATTTGGGCTATGAAA 147
DB 608720 CGCTTAGCTTTACAGAGAGGATGAAAGGCCCAATAATGATTGTAATGTAATGTA 608779

QY 148 GCGGCGCCTTGAGCGTTGCTGAAGACGCACTCGGAAACATCATCGGCCGCACTGAAGC 207
DB 608780 GAATATGATTTGCTATTCTGTCAGATTCATTTGGAATCTTTTATTCTGAAGGCAGGT 608839

QY 208 ACTGATCCGGAGCTTCTCGATCGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 267
DB 608840 AAGAAGATTTTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 608899

QY 268 GGGATGTTGATGGCACTGCGAGCGTGGTGGCGCCCTTGAGGCTGCGCGGTGATGCTG 327
DB 608900 GGTAAATTTGATGGCTTTGGGATCTGTTGCTGGGTGGAATCTTTTGGCAATTTGTT 608959

QY 328 GAGAACGGCTACGTGAATCGGCATCCATTTGATTTGATTCGCGATCGTGGAGGAGGAGG 387
DB 608960 GAACAGAAATTTGAGACTGCTTATCTCTTTAGAAATGTAATTTTACTTTGTAAGAAATCG 609019

QY 388 GCGCGCTTACGAGTGGCATGTTGGGCGCGCGGCGCATTCAGGGTTGGTCGCGCAGG 447
DB 609020 AGTAGATTTAAATTTGCTACATTTGGGTAGTAAAGTTATGTTGTCGCATAGTAATCAAGAA 609079

QY 448 GAACTGGACTCTTTGTTGATAGGATGAGTGTGCTGCTAGGCGGCGGCTACTGCTTC 507
DB 609080 AATTAAGTTTATGCTGATTAACAGGAAAGGTTTATCAGAACGTATGCTGCTGAAGTA 609139

QY 508 GCGTTGAACCGCGGCAACTGCAAGGTGCAAGCGGCTCGCGCGCGGCACTGCGTGTCTTT 567
DB 609140 GGAATGAAATTTTAAATTTGTTTAAATCAAGCAAAACGTTGATGCAAAAGGAAATTTAAATGTTTT 609199

QY 568 ATCGAATCAATTAACAGAGACGATCCTCGAGCAGGAGCAATAGAGATCGAGATT 627
DB 609200 TTTGAACTTATATAGAACAGAGCGCTCGTTTATAGAAATGAAAGGAAACAAATAGGTGTT 609259

QY 628 GTAACTTCAATCGTGGCGTTTCGCGATTCGCGATTCGCGTTGCGGTTCGCGTCAAGAGCAGACCGCAC 687
DB 609260 GTACAGGATTTGCTGCGCGGATTCGTGCAATTTGTTAAATTAAGACAGACGCGATCAT 609319

QY 688 GCCGGCACAAACCCCATGCACTTCGCGCAGGATGCGTGTGTACCGCGCTCTCATGGTG 747

DB 609320 TCAGGAGCAACCAATGCAATTCATGCTATGATGCTATGTTAGGAGGATCTGAGTTATCA 609379

QY 748 AGGGAGGTCAACCGGTTCTGTCACAGAGATCGCGATGGGCACAGTGGCTACCGTTGGCCAC 807

DB 609380 CTTGCTATTGAGGAGCTGCTATTCAAGCTGGACATTTACAGTAGCTACAGTAGTAAAT 609439

QY 808 CTCACAGTGGCCCCCGGTGGAGGCAACAGGTCGCCGGGGAGGTGGACTTTCACACTGGAC 867

DB 609440 ATTACAGCTAAACCCAGGAGTAATGAAATGTTGTGCCAGGATATTGCCAAATTTATTAGTAGAT 609499

QY 868 CTGCGTTCTCGCATGAGAGTGGCTCGCGGTGCTGATCGACCGCATCTCGGTATGGTTC 927

DB 609500 ATTGAGGTACACATGTCACAGCTAGAGATTCTGTATTTGAATTTATTACAAGAGGAAAT 609559

QY 928 GCGAGGTGCGCTCCCGAGCGGTGGCTGCGGATGTCGATGAAATTTTCAATCTCAGC 987

DB 609560 AGTAAAGTTTCAGAAAAGAGGATTTAATTAATGAGTTTACAACTTATTTCAAAAGATAAT 609619

QY 988 CCGGTGAGCTGGCTCTTACCATGTTGAGCGCGCTTCGCGAAGCGGCTCGGCTTTGCAG 1047

DB 609620 CCAATAATATTACCTGAAATATGTTAAACCAATAGCTGAAACACACATTCCTTGGT 609679

QY 1048 TTCACACCGGATATCAGAGTGGGGCGGCCACGACTCGATGTTTCATCGCCCAAGTTC 1107

DB 609680 TATTCTTACGAAATTAATGCAAGTGGTGCAGGACATGATGCAATGCAATGCAACACTT 609739

QY 1108 ACGGAGCTCGGAATGTTTTCGTTCCAAAGCGTGGCGGAGCCAGTTCCTCCGAAGAA 1167

DB 609740 TGCCCAACAGGTATGATTTTATTTCCATCTCATTTGGGAAATAGTCAATACCTCTTGAA 609799

QY 1168 TGACCGATTTCGATGACCTTCGCAAGGAACATGAGGTTGCTCTCGGGTAAATGAAG 1225

DB 609800 TTTACTGATTGGAAGATATAGAGCGAGGAATTAAGTTTACAAAAGTTATATTGG 609857

RESULT 9

US-09-489-039A-4854
; Sequence 4854, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4854
; LENGTH: 924
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4854

Query Match 8.5%; Score 104.8; DB 4; Length 924;
Best Local Similarity 47.5%; Pred. No. 1.1e-17;
Matches 310; Conservative 0; Mismatches 342; Indels 0; Gaps 0;

QY 569 TCGAACTACACATTGAACAGGACCGATCCTCGAGCAGGAGCAAAATAGAGATCGAGTTG 628

DB 245 TTGAGCTGATATCGAACAGGAGCAATTTCTTGATAAAGAGCAGATTGATATTGGCGTGG 304

QY 629 TAACTTCCATCGTTGGCGTTTCGGGCTTCGGGTTTCGGGTTTCGGGTTTCGGGTTTCGGG 688

DB 305 TTACCGCGGTACAGGATTTTCGTGGCAGGAAATTTACCTCGAGAGCGGTATCGAATCATG 364

QY 689 CCGGCACAAACCCCATGCACTGCGCCAGGATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTG 748

DB 365 CCGGTTACTACCAATGTCATGAGACGTGATCGCGGCTGCGGCTGCGGCTGCGGAAATCGTG 424

QY 749 GGGAGGTCAACCGGTTTCGTCAACAGATGCGCGATGCGCAGTGGGTACCGTTGGCCACC 808

```
Db 425 TTTTCCCGTGGAGTGGCGCTAAGCCTTGGTGAATCAGGTAGTACCGTTGGACATT 484
QY 809 TCACAGTGGCCCGCGGTGAGAGCAACAGGTCCCGGGGAGGTGAGACTTCACACTGGACC 868
Db 485 TCAGCGTGAAGCGTAATCTCTGATCAACGTCAATTCGCCAACCATGTCTGTCATGTCAAGTGGATC 544
QY 869 TCGGTTCTCCGATGAGGAGTCCGTCGGGTGCTGATCCAGCCGATCTCGGTCAATGGTCG 928
Db 545 TCCGCAATACCGAATATGGATCTTATGTCGCGGAGCAGCAATGGCCGAATTTGTGCG 604
QY 929 GCGAGTCCGCTCCCAAGCGCGTGTGGCTGCCGATGTGGATGAATTTTCAATCTCAGCC 988
Db 605 CGAARACGTGCGAGGAAGGGGTGAGATAACCAAGCCGCTCACTGGTGGCTTTAATC 664
QY 989 CGGTGAGCTGGCTCTCATTGAGTGGACCGCGTTCGCGAAGCGGCTCGGCTGGCTTGCAGT 1048
Db 665 CGGTCACTTTTCCGATGAGATCGTCAATCCGCTGGAAGCCGAGGCGGAGCGTCAAGGCGC 724
QY 1049 TCACACACCGGATATCAGACGTGGGCGGGCCAGACTCGATGTTTCATCGCCAGGTCA 1108
Db 725 TCAGTTACCGAGTTTGGCAAGCGGCGCCGACACATGCAATTTATGGCGTGGGTAT 784
QY 1109 CGACGCTCGGAATGTTTTGTTTCCAAAGCGGTGCTGGCCGGAGCCACAGTTTCCCGGAAGAT 1168
Db 785 GTCCCGCGGGATGATTTTGTCCCTGCTGATGGCATTAGCCATAACGTCAAGAATC 844
QY 1169 GGACCGATTTCGATGACCTTCGCAAGGAACGTGAGTTGCTCCCGGGTAAT 1220
Db 845 ATAGCGCAGCAAGATCTGATTTGCTGGCGCAATGTGCTGCGAGGTAGT 896

RESULT 10
US-09-543-681A-1026
; Sequence 1026, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; PRIOR FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 1026
; LENGTH: 1290
; TYPE: DNA
; ORGANISM: Proteus mirabilis
US-09-543-681A-1026

Query Match 7.8%; Score 97; DB 4; Length 1290;
Best Local Similarity 44.2%; Pred. No. 1.2e-15;
Matches 498; Conservative 0; Mismatches 620; Indels 9; Gaps 2;

QY 79 GGTGTTACCGGCTGACCTACACTCCAGAGATGCGCGCGCGGGAACGTCATTGCG 138
Db 130 GGGTAACACGTATTGGCTTATAGCGAAGAAGATGAAGCAGCCCATCTCTATCTTGTAGT 189
QY 139 GCTATGAAGCGCGCCCTTGAGCGTTTCGTGAAGACGCACTCGGAACATCATCGGCCGA 198
Db 190 GTAATGAAGAGCGAGCCCTTGAAGTCTATCTGACGGTATTGGCACACTCTACGCTGCG 249
QY 199 CGTGAAGGCATGATCCGAGCTTCTGCGATGCGGGTCCGTTCACTTCGATTCGTGTC 258
Db 250 TTACCGGGCAAGATAGAACATTAACCGCGGTAGGAACAGGCTCTCACCTTGATACCGTT 309
QY 259 CGAAGCGCGGATGTTTCATGCGACTGAGCGGCTGGTGTGCGCCCTTGAGGCTCCCGG 318
Db 310 CCCCAGGGGGCGCTTATATGCGCTTAGGAGTTATGCTGGTTTATGCTT----- 364
QY 319 GTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTGAGTTTCACTCGCGATCGTGGAG 378
```

```
Db 365 -TGATGCAATACAAACCACTGAACCTGATCTAGAGTTAATTTGTTATTCGAGCA 423
QY 379 GAGGAAGGGGCGCGCTTCAGCAGTGGCATGTTGGCGCGCGGCCCATTCGACGGGTGTC 438
Db 424 GAAGAATCGAGCCGATTTGGGTTTTTCATGTATCGGAAGTAAAGTGTCTAACAGGGAAT 483
QY 439 GCCGACAGGGAACCTGACTCTTTGTTGATGAGGATGAGTGTCCGTTAGGCGAGCGGCT 498
Db 484 GACCGTACTCGATGGGAACAAATAGAGATGATGAGGTAAATAATTTCTTTGATGTTTA 543
QY 499 ACTGCTTTCCGTTGAAGCCGCGGAACTGCAGGCTGCAGCCCGCTCCGCGCGGACCTG 558
Db 544 AAATCCCTTGGTTATCAGCATGAAAACCTTAGATCAATGTTTACTTTAAATCAGACCGTTAT 603
QY 559 CGTGTCTTTATCGAACTACATTTGAACAAGGACCGATCTTCGACGACGAGGCAATAGAG 618
Db 604 AGTGCAATTTGTTGAATTAATTAATTTGAACAAGGTAAACGCTTAGAATAACGATCAAAAACA 663
QY 619 ATCGGAGTTGTAAACCTCCATCGTTGGCGTTTCGCGCATTTGCGGTTGCGCTCAAGGCGAGA 678
Db 664 ATGGAATTTGCAATGTTATCGTGGCAACCGCAATGTTCAACGTTCAAGATGCGTTAGTGGCGAGTGG 723
QY 679 AGCGACACGCGCGGCAACACCCCATGCACTGCGCCAGGATGCGCTGTTACCCCGCGCT 738
Db 724 GCAGACCACTCTGGTGCAACCGCAATGTTCAACGTTCAAGATGCGTTAGTGGCGAGTGG 783
QY 739 CTATGTTGAGGAGGTCAACGGGTTGCTC---AACGAGATCGCGATGGGCACAGTGGCT 795
Db 784 GGTATTTACTGATTAATCAATCATGCTGCTCTGTAACGGAAGCGGTATACGGCACCGTAGGT 843
QY 796 ACCGTTGGCCACCTCACAGTGGCCCGCGTGGAGCAACAGGTCCCGGGGAGGTGGAC 855
Db 844 ACCATTGTTAACTTAATGTCATACCAAACTCGATGAACGTTATTCAGGCGCAAGTAAAA 903
QY 856 TTCACTACGTGACCTGCGTTCTCCGATGAGGAGTGGCTCCGCGTCTGATCGACCGCATC 915
Db 904 TTCTCCGGTGAATTTCCGTGGTATTGATACGGACACAGTATTCAACGTTGTTACAGCGTTTA 963
QY 916 TCGGTGATGTTGCGGAGGTGCGCTCCAGCGCGGTGCGCTGCGGATGGATGAATTT 975
Db 964 ACAACAGTGTGAGAAAGCAGAAAGATTTTGGTGTAAAAAATAATAGTGAACCTATT 1023
QY 976 TTCAATCTCAGCGGTGCGTCTCTACCTACCTGTTGGAGCGCGCTTCGGAAGCGGCC 1035
Db 1024 TCGGCTGAATCACCAGTAAACTGGATGACTCTATTTCGCAAGTATTGAAAGCATTTGT 1083
QY 1036 TCGGCTTTGAGTTTCAACACCGGGATATCAGCAGTGGGGCGGGCCACGACTCGATGTTTC 1095
Db 1084 CAAAAACACGATATCAATTAATGACTATGTTAAAGTGGTGGCGGTCTAGCTCAATGAAT 1143
QY 1096 ATCGCCAGGTTCAGGAGCTCGGAATGGTTTTCGTTCCAGCCGCTGCTGGCGGAGCCAC 1155
Db 1144 ATGGCATCTTTATACCCGACGAGCATGATTTTTTACCCCTCTGTTGGGGGATTAGTCAT 1203
QY 1156 GTTCCGGAAGATGAGCCGATTTTCGATGACCTTCGCAAGGAACCTGA 1202
Db 1204 CACCAGATGAATTTACGGAGTTTAGTGATATTGCAATAGCTGCAGA 1250
```

```
RESULT 11
US-09-489-039A-2592/c
; Sequence 2592, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; PRIOR FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
```

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 2592

; LENGTH: 480

; TYPE: DNA

; ORGANISM: Klebsiella pneumoniae

US-09-489-039A-2592

Query Match

Best Local Similarity 7.5%; Score 93.4; DB 4; Length 480;

Mismatches 238; Conservative 0; Mismatches 241; Indels 0; Gaps 0;

668 TCAAGCGCAGACCGACCGCGGCACACACCCCATGACCTGGCGCAGGATCGCGTGG 727

480 TTACCGCGGAGGTGGCCATGCGCGCACGGTCCGATGCTTCATCGCAAGGACGCCCTCG 421

728 TACCCGCGGCTCATGTGGTGGAGGAGTCAACCGGTTCGTCAACGAGATCGCCGATGGCA 787

420 CCGCGCGGCGGATGGATGGTGCAGGTGGAACCTGACCCGCGCGGCGGTAAAC 361

788 CAGTGGCTACCGTTGGCCACCTCACAGTGCCGCCCGGTGGAGCAACCAAGTCCCGGGG 847

360 TGTGGCGACGGTCGCGCACCTCGGCTGTGCGCGCGCGGGTAAACGTGATACCGGGCG 301

848 AGGTGCACTTCACTGGACCTCGGTTCTCCGATGAGGAGTCCGCGGTGCTGATCG 907

300 AGGTGAGTTGACGCTGGATATTTCGCGCGCCGCGAGGATGCGCGCTGACGCGCGCTGCTG 241

908 ACCGATCTCGGTCAATGTCGGGAGGTGCGCTCCCGAGCGCGGTGTGGCTGCGGATGTG 967

240 AGGAATCTGCGGCGAGCGCAGCGATGTCGCGGCGCGCCGACGTGATTTCCGCGCG 181

968 ATGAATTTTCAATCTCAGCCCGGTGAGCTGCTCTACCATGTGGAGCGCGTTTCGCG 1027

180 AGGATATTATCATCTCGCGCGCACCGCTGACAGTCACTGCGAGCGGTGCTGAGCG 121

1028 RAGCGCCTCGGCTTCAGTTTACACACCGGATATCAGCATGGCGGCGGCACGACT 1087

120 AGCGGTGTGGCGGTGCGAGGCGGTTCCTGACGTGCGGCGCGCAGCGCATGACG 61

1088 CGATGTTTATCGCGCCAGGTCAACGAGTCCGAAATGTTTCGTTTCAAGCGGTGCTGGC 1146

60 CCATCCCATCGCGGAGCGGTGCCATCGCCATGCTGTTGTACGTTGCTGGGGGC 2

RESULT 12

US-09-252-991A-6967/c

; Sequence 6967, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6967

; LENGTH: 525

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6967

Query Match

Best Local Similarity 7.5%; Score 93.2; DB 4; Length 525;

Mismatches 209; Conservative 0; Mismatches 193; Indels 0; Gaps 0;

820 CCGGTGGAGGCAACAGGTCCCGGGGAGGTGGATTCACACTGCACTGCGTTCTCG 879

521 CCGGTTCGCGCAACGTGATACCCGCGAAGTGAAGATGACCCCTGGACTTCGCGCATCTG 462

880 CATGAGGAGTCGCTCGCGTGTGCTGATCGACCGCATCTCGGTCACTGTGTCGCGAGGTGCC 939

461 CAACCGGAGCGCTGACTCGATGATCGCGAAGTCCGCAAGTATCGCCGCTACTGTC 402

940 TCCAGGCGGTGTGCTGCGGATGTGATGAATTTTCAATCTCAGCGCGGTGCGAGCTG 999

401 GAGAAGCATGGCTTGAATACGAGCTGGTCCGCGCGGATTTCCGCGCGTGTACTTC 342

1000 GCTCCTTACCATGTGTGACGCGTTCGCGAAGCGGCTCGGCTTGCAGTTTCAACACCGG 1059

341 GACCAGGATGCTGCGCGGTGCGCGAGCGCGCGAGCGCTGGGCAATGCGCGAGATG 282

1060 GATATCAGCAGTGGGCGGCGCACGACTGATGTTTTCATGCGCCAGGTTCACGAGCTCGGA 1119

281 GACATGTCAGCGGCGCGCGCACGCGATCTTCTCGCGAATCTCGTCCGCGGG 222

1120 ATGCTTTTGTTCGAAGCGGTGTCGCGGAGCGCATGTTCCCGAAGATGAGACCGATTC 1179

221 ATGATCTTGTGCTGCGAGAACGCGCATCAGCCACAGAGATCGAGAACGCCAGCCCC 162

1180 GATGACCTTCGCAAGGAACTGAGGTTGCTCCCGGGTAATG 1221

161 GAGACCTGCGCGCGGCTGCGCGGTGCTGTTGCGGGGCAATG 120

RESULT 13

US-09-252-991A-6968/c

; Sequence 6968, Application US/09252991A

; Patent No. 6551795

; GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6968

; LENGTH: 564

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6968

Query Match

Best Local Similarity 6.8%; Score 84; DB 4; Length 564;

Mismatches 189; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

146 AAGCGCGCGCTTGACGCTTCTGTAAGACGCACTCGGAAACATCATCGGCGCGGAGTGAAG 205

364 AAGCGCGCGCTGCGACGCTGCGGTGCGGTGCGGCAACATCTTCGCGCGCTCGTCCCG 305

206 GCACTGATCCGAGCTTCTCGCGATCGCGGTTCGCTTTCACACTTCGATTCGTCCGAAACG 265

304 GCGGTAAACCGGACTTTCGCGCGGTGATGACCGGTAGCCACATCGACACCCAGCCACCG 245

266 GCGGATGTTTGTGCGACTGCGAGCGGTGTGCGCGCTTCGAGGCTGCGCGGCTGATGC 325

244 GTGGCAAGTTTCGACGCTTCTCGGGTGTATGCGCGGCTCGAGGTGATCCGACCCCTCA 185

326 TCGAGAACGCTTACGTGAATCGCATTCATTTGATTCATCGCGATCTGCTGGAGGAGGAAG 385

184 ACGACCTCGGGGTGGAACCGAGGCGCGCTGCGAGGTGTGTGGACCAACAGAGGAAG 125

386 GCGCGCGCTTACGAGTGGCATGTTGGGCGCGCGGCCATTTCAGGTTGCTGCGCGACA 445

124 GCTCGGCTTTCGCGCGCTGATGAGGCTCGGGGTATTCGCGGGAAGTTTACCCTGG 65

446 GGGAACTGGACTCTTTTGTGTTGATGAGGATGGAGTGTCCGTTAGGCGGCGGCTACTGCT 505

Db 64 AGGAGACCTGGCCAGCGGATGCCGACGGTGTACAGCTAGGCGAGGCGCTGGAGCCCA 5
QY 506 TCGG 509
Db 4 TCGG 1

RESULT 14

US-09-489-039A-2583/c
; Sequence 2583, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 2583
; LENGTH: 339
; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-2583

Query Match 6.2%; Score 76.8; DB 4; Length 339;
Best Local Similarity 52.1%; Pred. No. 1.3e-10;
Matches 171; Conservative 0; Mismatches 157; Indels 0; Gaps 0;

QY 73 GSCCCCGGTGTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGCGGAAACGCTC 132
Db 328 GCCGACGCCCTGACCCGGGTATCTGTCTGCCGAGCATCTGCAGGCCAACACGCTGGTG 269
QY 133 ATTGCGGCTATGAAGCGCGCGCTTGAGCGTTGCTGGAAGACGCACTCGGAAACATCATC 192
Db 268 GGGCAGTGTATGAGGCTGCCGGCATGATGGTGTGGCAGCAGCGTTCGGCAACATCTGC 209
QY 193 GSCCGACGTGAAGGCACTGATCCGAGCTTCTTCGCGATCGCGGTTCGTTTCACTTCGAT 252
Db 208 CGACGTTATGAAGCAACAGGAGGGCGCGCGGTCTGCTGGGCTCTCATCTTGAC 149
QY 253 TCTGTCCGAAGCGGGGATGTTGATGGCACTGCAGGCGGTGTGCGCCCTTGAGGCT 312
Db 148 ACCGTGCGCAACGCCGACGCTACGACGCATGCTCGCGGTCTGCGCGCGATTGAGGTG 89
QY 313 GCCCGGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCCATTTGAGTTTCATCGCATC 372
Db 88 GTCCAGCTCTGCACCGCATGGCGACGGCTGGCGAAGGCAATCGAGATCGTTCGGCTTT 29
QY 373 GTGGAGGAGGAAGGGCCCGCTTCAGCA 400
Db 28 CGCGATGAAGAGGCGCACCCGCTTTGGCA 1

RESULT 15

US-09-489-039A-4973
; Sequence 4973, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
; FILE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 4973
; LENGTH: 405

; TYPE: DNA
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-4973
Query Match 4.6%; Score 56.6; DB 4; Length 405;
Best Local Similarity 55.2%; Pred. No. 2.1e-05;
Matches 132; Conservative 0; Mismatches 104; Indels 3; Gaps 1;
QY 72 AGGCCCCGGTGTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGCGGAAACGCT 131
Db 102 AGGCGCGCGGTATGCCGCCCTGGCAGCAGCTGCCGAAGATAAAGCCGCGAGGACTTTGT 161
QY 132 CATTCGGCTATGAAAGCGCGCCCTTCAGCGTTTCGTGAAGACGCACTCGGAAACATCAT 191
Db 162 TGTAGCCCGCATGAAGCGCTGGGCTGAACGTGTCAATTGATGCTATCGGGAATGTCAC 221
QY 192 CGGCGCAGCTGAAGGCACTGATCCGAGAGCTTCCTGCGATCCGCGGTTCACACTTGA 251
Db 222 CGGCGTTTATCACGCGCAGGAGACGTTGCCGATGGTGAT---GATGGGCTCACATATTGA 278
QY 252 TTCTGTCCGAACGCGGGGATGTTTGTATGGCACTGCAGGCGTGTGTGCGCCCTTGAGS 310
Db 279 TACGTTCCGCCACCGATGGGTTATACGATGGCAAAATACGCGCGTTATGGCCGCTTGAGS 337

Search completed: May 4, 2004, 16:04:11
Job time : 141 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 4, 2004, 15:00:24 ; Search time 618 Seconds
(without alignments)
9068.887 Million cell updates/sec

Title: US-10-045-063-1
Perfect score: 1239
Sequence: 1 G9gacctgcagaagcgca.....tgaaggcacttgaccggttaa 1239

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2936184 seqs, 2261732022 residues

Total number of hits satisfying chosen parameters: 5872368

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications NA:
1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq:
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq:
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq:
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq:
6: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq:
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq:
8: /cgn2_6/ptodata/2/pubpna/US08_PUBCOMB.seq:
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq:
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq:
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq:
13: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq:
14: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq:
15: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq:
16: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq:
17: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq:
18: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq:
19: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1238	99.9	1239	9	US-09-916-501-3
2	1238	99.9	1239	15	US-10-334-990-6
3	1192.4	96.2	1263	9	US-09-916-501-6
4	739.8	59.7	1239	9	US-09-950-772-5
5	739.8	59.7	1239	15	US-10-289-360-3
6	739.8	59.7	3343	9	US-09-950-772-7
7	739.8	59.7	3343	15	US-10-289-360-5
8	220.6	17.8	1163020	16	US-10-398-221-10
9	220.6	17.8	3011208	16	US-10-398-221-2058
c 10	158.6	12.8	3880	16	US-10-398-221-3817
11	150.2	12.1	1218	9	US-09-974-300-308
12	122.6	9.9	32768	9	US-09-070-927A-123
13	121.2	9.8	1830121	15	US-10-329-960-1
14	121.2	9.8	1830121	16	US-10-329-670-1
					Sequence 3, Appli
					Sequence 6, Appli
					Sequence 5, Appli
					Sequence 3, Appli
					Sequence 7, Appli
					Sequence 5, Appli
					Sequence 10, Appli
					Sequence 2058, Ap
					Sequence 3817, Ap
					Sequence 308, App
					Sequence 123, App
					Sequence 1, Appli
					Sequence 1, Appli

15	112.8	9.1	1239	15	US-10-156-761-3482	Sequence 3482, Ap	
16	112.8	9.1	9025608	15	US-10-156-761-1	Sequence 1, Appli	
17	111.6	9.0	582	9	US-09-974-300-4815	Sequence 4815, Ap	
18	88.4	7.1	1657	13	US-10-425-114-11825	Sequence 11825, A	
19	86.8	7.0	1218	9	US-09-974-300-297	Sequence 297, App	
20	86.6	7.0	1772	13	US-10-424-599-87117	Sequence 87117, A	
21	82.8	6.7	4525	13	US-10-424-599-87115	Sequence 87115, A	
22	75.8	6.1	1731	13	US-10-425-114-7758	Sequence 7758, Ap	
23	65	5.2	693	9	US-09-974-300-4808	Sequence 4808, Ap	
24	60.2	4.9	2262	13	US-10-424-599-15592	Sequence 15592, A	
C	25	57.6	4.6	466	9	US-09-070-927A-943	Sequence 943, App
26	54	4.4	648	13	US-10-282-132A-11544	Sequence 11544, A	
27	53	4.3	1140	16	US-10-369-493-34696	Sequence 34696, A	
28	50.8	4.1	1215	15	US-10-156-761-1598	Sequence 1598, Ap	
29	49.2	4.0	1131	16	US-10-369-493-42864	Sequence 42864, A	
30	48.6	3.9	834	15	US-10-156-761-2613	Sequence 2613, Ap	
31	48.6	3.9	1716	9	US-09-815-242-7716	Sequence 7716, Ap	
32	48.6	3.9	1716	13	US-10-282-132A-30050	Sequence 30050, A	
33	48.4	3.9	2205	16	US-10-369-493-31916	Sequence 31916, A	
34	47.4	3.8	12441	10	US-09-988-384B-3	Sequence 3, Appli	
35	47.4	3.8	13613	9	US-09-861-289-3	Sequence 3, Appli	
36	47.4	3.8	13613	9	US-09-860-846-3	Sequence 3, Appli	
37	47.4	3.8	13613	10	US-09-836-821-3	Sequence 3, Appli	
38	47.4	3.8	13613	15	US-10-271-889-46	Sequence 46, Appl	
39	47.4	3.8	38506	10	US-09-793-708-19	Sequence 19, Appl	
40	47.4	3.8	38506	15	US-10-201-365-1	Sequence 1, Appli	
41	47.4	3.8	38506	15	US-10-160-539-19	Sequence 19, Appl	
42	47.2	3.8	1958	16	US-10-108-260A-1003	Sequence 1003, Ap	
43	46.2	3.7	3504	16	US-10-369-493-35616	Sequence 35616, A	
C	44	45.6	3.7	609	15	US-10-156-761-5800	Sequence 5800, Ap
45	45.6	3.7	1725	15	US-10-156-761-7096	Sequence 7096, Ap	

ALIGNMENTS

RESULT 1

US-09-916-501-3
; Sequence 3, Application US/09916501
; Patent No. US20020132848A1
; GENERAL INFORMATION:
; APPLICANT: KRIMMER, Hans-Peter
; APPLICANT: REICHERT, Dietmar
; APPLICANT: DRAUZ, Karlheinz
; APPLICANT: KLEMENT, Ingo
; APPLICANT: MAY, Oliver
; TITLE OF INVENTION: Process for the Preparation of Allylsine Acetal
; FILE REFERENCE: 210740US-10757-9350-0-X
; CURRENT APPLICATION NUMBER: US/09/916,501
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: Germany 100 37 115.9
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 3
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Arthrobacter aureusens
US-09-916-501-3

Query Match	99.9%	Score 1238;	DB 9;	Length 1239;
Best Local Similarity	100.0%;	Pred. No. 0;		
Matches 1238;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	2	TGACCTTCGAGAACGCGCGCGCATTCGAGAACGATCCCGGAGCTCCCGGT	61	
Db	2	TGACCTTCGAGAACGCGCGCGCATTCGAGAACGATCCCGGAGCTCCCGGT	61	
Qy	62	TCTCGGACAGAGCCCGCGGTGTACCCGCGTACCTACACTCCAGAGCATGCCGCGCGC	121	
Db	62	TCTCGGACAGAGCCCGCGGTGTACCCGCGTACCTACACTCCAGAGCATGCCGCGCGC	121	
Qy	122	GGGAAACGCTCATTCGCGGTATGAAAGCGCGCCCTTGAGCGTTGCTGAAGACGCACTCG	181	

Db 122 GGGAAACGCTCATTTGGGCTATGAAGCGCGCCGCTTGAGCGTTCTGTGAAGACGCACTCG 181
QY 182 GAAACATCATCGGCCGACGCTGAAGCACTGATCCGAGCTTCTCGCATCGCGGTGCGTT 241
Db 182 GAAACATCATCGGCCGACGCTGAAGCACTGATCCGAGCTTCTCGCATCGCGGTGCGTT 241
QY 242 CACACTTCGATTCTGTCCGAAACGCGCGGATGTTTGAATGGCACTGCAAGCGGTGTTGCG 301
Db 242 CACACTTCGATTCTGTCCGAAACGCGCGGATGTTTGAATGGCACTGCAAGCGGTGTTGCG 301
QY 302 CCCTTGAGGCTCGCGCGGTGATCTGGAGAACGCGCTACGTAATCGGCATCCATTTTAGT 361
Db 302 CCCTTGAGGCTCGCGCGGTGATCTGGAGAACGCGCTACGTAATCGGCATCCATTTTAGT 361
QY 362 TCATCGCATCTGTGAGGAGGAGGCGCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
Db 362 TCATCGCATCTGTGAGGAGGAGGCGCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
QY 422 CCATTGACGGTGTGTCGCGACAGGGAACCTGGAATCTTTGTTGATGAGGATGAGTGT 481
Db 422 CCATTGACGGTGTGTCGCGACAGGGAACCTGGAATCTTTGTTGATGAGGATGAGTGT 481
QY 482 CCGTTAGGCGGCTACTGCTTCGCGCTTGAAGCGCGGCAACTGCAGGCTGCAGCCC 541
Db 482 CCGTTAGGCGGCTACTGCTTCGCGCTTGAAGCGCGGCAACTGCAGGCTGCAGCCC 541
QY 542 GCTCCGCGCGGACCTGCTGCTTTTATCGAACTACACATTGAACAGGACCGATCCTCG 601
Db 542 GCTCCGCGCGGACCTGCTGCTTTTATCGAACTACACATTGAACAGGACCGATCCTCG 601
QY 602 AGCAGAGCAATAGAGATCGAGTTGTAACCTTCATCGTTCGCGGCTTCGCGATTGCGGG 661
Db 602 AGCAGAGCAATAGAGATCGAGTTGTAACCTTCATCGTTCGCGGCTTCGCGATTGCGGG 661
QY 662 TTGCGCTCAAGGCAAGGACCAACGCGGCAACCCCATGACCTGCGCCAGGATG 721
Db 662 TTGCGCTCAAGGCAAGGACCAACGCGGCAACCCCATGACCTGCGCCAGGATG 721
QY 722 CGTGTGTACCCGCGCTCTCATGTTGAGGAGGTCAACCGGTTTGTCAACGAGATGCGCG 781
Db 722 CGTGTGTACCCGCGCTCTCATGTTGAGGAGGTCAACCGGTTTGTCAACGAGATGCGCG 781
QY 782 ATGGCACAGTGGTACCGTTGGCCACCTCAGATGGCCCCCGGTGGAGCAACCGATCC 841
Db 782 ATGGCACAGTGGTACCGTTGGCCACCTCAGATGGCCCCCGGTGGAGCAACCGATCC 841
QY 842 CGGGGAGGTGGACTTTCACACTGGACCTCGTTCTCCGCATGAGGAGTCCGTCGCGGTGC 901
Db 842 CGGGGAGGTGGACTTTCACACTGGACCTCGTTCTCCGCATGAGGAGTCCGTCGCGGTGC 901
QY 902 TGATCGACCGCATCTCGGTTCATGTCGCGGAGGTGCGCTCCAGGCGGTGTGGTGC 961
Db 902 TGATCGACCGCATCTCGGTTCATGTCGCGGAGGTGCGCTCCAGGCGGTGTGGTGC 961
QY 962 ATGTGATGAAATTTTCAATCTCAGCCCGGTGAGTGGCTCTTACCATGTTGGTGC 1021
Db 962 ATGTGATGAAATTTTCAATCTCAGCCCGGTGAGTGGCTCTTACCATGTTGGTGC 1021
QY 1022 TTTCGGAAGCGGCTCGGCTTTCAGTTTCACACCGGATATCAGAGTGGGGGGG 1081
Db 1022 TTTCGGAAGCGGCTCGGCTTTCAGTTTCACACCGGATATCAGAGTGGGGGGG 1081
QY 1082 ACAGCTCGATGTTTCATCGCCAGGTCAAGGAGTGGGATGTTTTCGTTCAAGCGGTG 1141
Db 1082 ACAGCTCGATGTTTCATCGCCAGGTCAAGGAGTGGGATGTTTTCGTTCAAGCGGTG 1141
QY 1142 CTGGCGGAGGACCGTTCGGAAGATGACGATTTTCATGACCTTCGCAAGGAGT 1201
Db 1142 CTGGCGGAGGACCGTTCGGAAGATGACGATTTTCATGACCTTCGCAAGGAGT 1201
QY 1202 AGGTTGCTCTCGGGTAAATGAAGGCACCTTGACCGGTAA 1239

Db 1202 AGGTTGCTCTCGGGTAATGAAGGCACCTTGACCGGTAA 1239

RESULT 2
US-10-334-990-6
; Sequence 6, Application US/10334990
; Publication No. US20030175910A1
; GENERAL INFORMATION:
; APPLICANT: ALTENBUCHNER, JOSEF
; APPLICANT: BOMMARIUS, ANDREAS
; APPLICANT: MATTES, RALF
; APPLICANT: SYLDATK, CHRISTOPH
; APPLICANT: TISCHER, WILHELM
; APPLICANT: WIESE, ANJA
; APPLICANT: WILMS, BURKARD
; TITLE OF INVENTION: WHOLE CELL CATALYST
; FILE REFERENCE: 9350-0142-0
; CURRENT APPLICATION NUMBER: US/10/334,990
; CURRENT FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US/09/407,062
; PRIOR FILING DATE: 1999-09-28
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Arthrobacter aureescens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1239)
; OTHER INFORMATION:
US-10-334-990-6

Query Match 99.9%; Score 1238; DB 15; Length 1239;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1238; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 TGACCCCTCAGAAAGCGCAAGCGCGCGCATTTGAGAAAGATCCGGGAGCTCTCCCGT 61
Db 2 TGACCCCTCAGAAAGCGCAAGCGCGCGCATTTGAGAAAGATCCGGGAGCTCTCCCGT 61
QY 62 TCTCGGCAGAAAGCGCCCGGTGTTACCCGGCTGACCTACACTCCAGAGATGCGCGCGC 121
Db 62 TCTCGGCAGAAAGCGCCCGGTGTTACCCGGCTGACCTACACTCCAGAGATGCGCGCGC 121
QY 122 GGGAAACGCTCATTTGCGGTATGAAAGCGCGCCCTTGAGCGTTTGTGAAGACGCACTCG 181
Db 122 GGGAAACGCTCATTTGCGGTATGAAAGCGCGCCCTTGAGCGTTTGTGAAGACGCACTCG 181
QY 182 GAAACATCATCGCGCGCATGAAAGCACTGATCCGAGCTTCCTCGATCGCGGTGTT 241
Db 182 GAAACATCATCGCGCGCATGAAAGCACTGATCCGAGCTTCCTCGATCGCGGTGTT 241
QY 242 CACACTTCGATTCTCTCGAAACGCGCGGATGTTTGTATGGCACTGCAGGCGTGTGCG 301
Db 242 CACACTTCGATTCTCTCGAAACGCGCGGATGTTTGTATGGCACTGCAGGCGTGTGCG 301
QY 302 CCCTTGAGGCTCGCGCGGTGATCTGAGAGAACGGCTACGTAATCGGCATTCATTGAGT 361
Db 302 CCCTTGAGGCTCGCGCGGTGATCTGAGAGAACGGCTACGTAATCGGCATTCATTGAGT 361
QY 362 TCATCGCATCTGTGAGGAGGAGGCGCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
Db 362 TCATCGCATCTGTGAGGAGGAGGCGCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 421
QY 422 CCATTGACGGTGTGTCGCGCAACGAGGAACTGGAATCTTTGTTGATGAGGATGAGTGT 481
Db 422 CCATTGACGGTGTGTCGCGCAACGAGGAACTGGAATCTTTGTTGATGAGGATGAGTGT 481
QY 482 CCGTTAGGAGGCGGCTACTGCTTCGGCTTGAAGCGCGGCAACTGCAGGCTGCAGCCC 541
Db 482 CCGTTAGGAGGCGGCTACTGCTTCGGCTTGAAGCGCGGCAACTGCAGGCTGCAGCCC 541

QY 542 GCTCCGCGGACCTGCGTCTTTTATCGAACTACACATTGAAACAAGACCGATCCTCG 601
Db 542 GCTCCGCGGACCTGCGTCTTTTATCGAACTACACATTGAAACAAGACCGATCCTCG 601
QY 602 AGCAGAGCAATAGAGATCGAGTTGTAACCTCCATCTTGGCGTTCCGCAATTCGGGG 661
Db 602 AGCAGAGCAATAGAGATCGAGTTGTAACCTCCATCTTGGCGTTCCGCAATTCGGGG 661
QY 662 TTGCGCTCAAGGAGAGACGACGACCGCGGCAACAACCCCATGSCACTCTGCGCCAGATG 721
Db 662 TTGCGCTCAAGGAGAGACGACGACCGCGGCAACAACCCCATGSCACTCTGCGCCAGATG 721
QY 722 CGCTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCGG 781
Db 722 CGCTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCGG 781
QY 782 ATGGCAGATGGCTACCGTTGGCCACTCACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
Db 782 ATGGCAGATGGCTACCGTTGGCCACTCACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
QY 842 CGGGGAGGTGGACTTTCACATGAGACTCTGCTTCTCCGATGAGGAGTCTGCTCCGCGTGC 901
Db 842 CGGGGAGGTGGACTTTCACATGAGACTCTGCTTCTCCGATGAGGAGTCTGCTCCGCGTGC 901
QY 902 TGATCAGCGCATCTCGGTATGCTCGGAGGTCCCGAGGTCCTCCAGGCGGTGTGGCTGCCG 961
Db 902 TGATCAGCGCATCTCGGTATGCTCGGAGGTCCCGAGGTCCTCCAGGCGGTGTGGCTGCCG 961
QY 962 ATGTGATGAATTTTCAATCTCAGCCCGGTGAGCTGCTTACCATGTTGGAGCGCG 1021
Db 962 ATGTGATGAATTTTCAATCTCAGCCCGGTGAGCTGCTTACCATGTTGGAGCGCG 1021
QY 1022 TTGCGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGAGTGGGGCGGGCC 1081
Db 1022 TTGCGAAGCGGCTCGGCTTTGAGTTTACACACCGGGATATCAGAGTGGGGCGGGCC 1081
QY 1082 AGCACTCGATGTTTCGCGCCAGGTCAACGAGTGGGATGTTTCTTCAAGCGGTG 1141
Db 1082 AGCACTCGATGTTTCGCGCCAGGTCAACGAGTGGGATGTTTCTTCAAGCGGTG 1141
QY 1142 CTGGCGGAGGACGCTTCCGAGATGAGGACGATTCGATGACCTTCCAAAGGAACTG 1201
Db 1142 CTGGCGGAGGACGCTTCCGAGATGAGGACGATTCGATGACCTTCCAAAGGAACTG 1201
QY 1202 AGGTTGCTCCGGGTAATGAAGGCACTTGACCGGTAA 1239
Db 1202 AGGTTGCTCCGGGTAATGAAGGCACTTGACCGGTAA 1239

RESULT 3

US-09-916-501-6
; Sequence 6, Application US/09916501
; Patent No. US20020132848A1
; GENERAL INFORMATION:
; APPLICANT: KRIMMER, Hans-Peter
; APPLICANT: REICHERT, Dietmar
; APPLICANT: DRAUZ, Karlheinz
; APPLICANT: KLEMENT, Ingo
; APPLICANT: MAY, Oliver
; TITLE OF INVENTION: Process for the Preparation of Allylsine Acetal
; FILE REFERENCE: 210740US-10757-9350-0-X
; CURRENT APPLICATION NUMBER: US/09/916,501
; CURRENT FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: Germany 100 37 115.9
; PRIOR FILING DATE: 2000-07-28
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 6
; LENGTH: 1263
; TYPE: DNA
; ORGANISM: Arthrobacter aureus
; FEATURE:
; NAME/KEY: misc_feature

; LOCATION: (25)..(25)
; OTHER INFORMATION: n=any nucleotide
US-09-916-501-6
Query Match 96.2%; Score 1192.4; DB 9; Length 1263;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 1208; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
QY 2 TGACCTCGCAGAAAGCGCAGCGCGCATTTAGAAAGAGATCCGGGAGCTCTCCCGGT 61
Db 2 TGACCTCGCAGAAAGCGCAGCGCGCATTTAGAAAGAGATCTGGGAGCTCTCCCGGT 61
QY 62 TCTCGCAGAAAGCGCAGCGCGGTGTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGC 121
Db 62 TCTCGCAGAAAGCGCAGCGCGGTGTACCCGGCTGACCTACACTCCAGAGCATGCCCGCGC 121
QY 122 GGGAAACGCTCATTCGGGCTATGGAAGCGCGCGCTTGGAGCGTTCTGTAAGACGCTCTCG 181
Db 122 GGGAAACGCTCATTCGGGCTATGGAAGCGCGCGCTTGGAGCGTTCTGTAAGACGCTCTCG 181
QY 182 GAAACATCATTCGGCGGACGCTGAAGGCACTGATCCGGAGCTTCTTGGCATCGCGGTGGTT 241
Db 182 GAAACATCATTCGGCGGACGCTGAAGGCACTGATCCGGAGCTTCTTGGCATCGCGGTGGTT 241
QY 242 CACACTTCGATTTCTGTCGAAACGCGCGGATGTTGATCGCACTGCAGCGGTGGTGGCG 301
Db 242 CACACTTCGATTTCTGTCGAAACGCGCGGATGTTGATCGCACTGCAGCGGTGGTGGCG 301
QY 302 CCCTTTGAGGCTCCCGGGTGATGCTGGAGAAACGGCTACGTGAATCGGCATTCATTTGAGT 361
Db 302 CCCTTTGAGGCTCCCGGGTGATGCTGGAGAAACGGCTACGTGAATCGGCATTCATTTGAGT 361
QY 362 TCATCGCGATCTGAGGAGGAGGCGCGCTTTCAGCAGTGGCATGTTGGCGCGCGCGG 421
Db 362 TCATCGCGATCTGAGGAGGAGGCGCGCTTTCAGCAGTGGCATGTTGGCGCGCGCGG 421
QY 422 CCATTGAGGCTTGGTTCGCGGACAGGAACTGGACTCTTTTGTGATGAGGATGGAGTGT 481
Db 422 CCATTGAGGCTTGGTTCGCGGACAGGAACTGGACTCTTTTGTGATGAGGATGGAGTGT 481
QY 482 CGGTTAGGCGAGCGGCTACTGCTTCGGCTTGAAGCGCGGCGAATGCGAGGCTGACGCC 541
Db 482 CGGTTAGGCGAGCGGCTACTGCTTCGGCTTGAAGCGCGGCGAATGCGAGGCTGACGCC 541
QY 542 GCTCCGCGCGACCTGCGTCTTTTATCGAACTACACATTGAAACAAGACCGATCCTCG 601
Db 542 GCTCCGCGCGACCTGCGTCTTTTATCGAACTACACATTGAAACAAGACCGATCCTCG 601
QY 602 AGCAGAGCAATAGAGATCGAGTTGTAACTCCATCGTTGGCGTTCCGCGATTCGCGG 661
Db 602 AGCAGAGCAATAGAGATCGAGTTGTAACTCCATCGTTGGCGTTCCGCGATTCGCGG 661
QY 662 TTGCGCTCAAGGAGAGACGACGACCGCGGCAACAACCCCATGSCACTCTGCGCCAGATG 721
Db 662 TTGCGCTCAAGGAGAGACGACGACCGCGGCAACAACCCCATGSCACTCTGCGCCAGATG 721
QY 722 CGCTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCGG 781
Db 722 CGCTGTACCGCGCTCTCATGCTGAGGAGGTCAACCGGTTCTGTCACAGAGATCGCGG 781
QY 782 ATGGCAGATGGCTACCGTTGGCCACTCACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
Db 782 ATGGCAGATGGCTACCGTTGGCCACTCACAGTGGCCCGCGGTGGAGGCAACAGGTCC 841
QY 842 CGGGGAGGTGGACTTTCACATGAGACTCTGCTTCTCCGATGAGGAGTCTGCTCCGCGTGC 901
Db 842 CGGGGAGGTGGACTTTCACATGAGACTCTGCTTCTCCGATGAGGAGTCTGCTCCGCGTGC 901
QY 902 TGATCAGCGCATCTCGGTATGCTCGGAGGTCCCGAGGTCCTCCAGGCGGTGTGGCTGCCG 961
Db 902 TGATCAGCGCATCTCGGTATGCTCGGAGGTCCCGAGGTCCTCCAGGCGGTGTGGCTGCCG 961
QY 962 ATGTGATGAATTTTCAATCTCAGCCCGGTGAGCTGCTTACCATGTTGGAGCGCG 1021

||||| 962 ATGTGATGAATTTTTCATCTCAGCCCGGTGAGTGGCTCTTACCATGTTGGAGCGG 1021
Db
1022 TTTCGGAAGCGCTCGGCTTGCAGTTTACACACCGGATATCAGAGTGGGGCGGGCC 1081
Qy
1022 TTTCGGAAGCGCTCGGCTTGCAGTTTACACACCGGATATCAGAGTGGGGCGGGCC 1081
Db
1082 AGCACTCGATGTTTCATCGCCCGAGGTACCGAGTGGGAATGTTTTCGTTCCAAAGCCGTG 1141
Qy
1082 AGCACTCGATGTTTCATCGCCCGAGGTACCGAGTGGGAATGTTTTCGTTCCAAAGCCGTG 1141
Db
1142 CTGGCGGAGCCAGCTTCCGGAAGATGACCGAATTCGATGACCTTCCGAAAGGAATG 1201
Qy
1142 CTGGCGGAGCCAGCTTCCGGAAGATGACCGAATTCGATGACCTTCCGAAAGGAATG 1201
Db
1202 AGGTTGTCTCCGGGTAAATGAAGGCACTTGACCGG 1236
Qy
1202 AGGTTGTCTCCGGGTAAATGAAGGCACTTGACCGG 1236
Db
RESULT 4
US-09-950-772-5
; Sequence 5, Application US/09950772
; Patent No. US20020102713A1
; GENERAL INFORMATION:
; APPLICANT: SUZUKI, Shunichi
; APPLICANT: ONISHI, No. US20020102713A1Imasa
; APPLICANT: YOKOZAKI, Kenzo
; TITLE OF INVENTION: 5-SUBSTITUTED HYDANTOIN RACEMASE, DNA CODING FOR THE RACEMASE, AN
; FILE OF INVENTION: PROCESSES FOR PRODUCING OPTICALLY ACTIVE AMINO ACIDS
; FILE REFERENCE: 212318USO
; CURRENT APPLICATION NUMBER: US/09/950,772
; CURRENT FILING DATE: 2001-09-13
; PRIOR APPLICATION NUMBER: JP2000-278571
; PRIOR FILING DATE: 2000-09-13
; PRIOR APPLICATION NUMBER: JP2001-65815
; PRIOR FILING DATE: 2001-03-08
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 1239
; TYPE: DNA
; ORGANISM: Microbacterium liquefaciens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)..(1236)
; OTHER INFORMATION:
US-09-950-772-5
Query Match 59.7%; Score 739.8; DB 9; Length 1239;
Best Local Similarity 74.8%; Pred. No. 8.4e-212;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
Qy 1 GTGACCTCGACAAAGCGCAAGCGCGCATTTGAGAAAGATCCGGAGCTCTCCCGG 60
Db 1 GTGACCTCGACAGCGCGCGCGCATCGCATCGAGGAGACTCTGGACTCTCTCCCGC 60
Qy 61 TTCTCGGCAAGAGCCCGCGTGTATCCCGCTGACCTACACTCCAGAGATCGCCCGCG 120
Db 61 TTCTCGGTCGAAGGCGCGCGGTGACACGTCTCAGTACACTCCGAGACACGCCCGCG 120
Qy 121 CGGGAACGCTTATTCGCGCTATGAAGCGCGCGCTTTCGAGCGTTTCGTAAGACGCTC 180
Db 121 CGAGAGGTGATCGTCGCGCCCATGACGCGGAGCGGGGCTGAGCGTCCAGAGGACCTCTC 180
Qy 181 GGAACATCATCGCGCGAGCTGAAGGCACTGATCCGGAGCTTCTTCCGATCGCGTCTCGT 240
Db 181 GGAACATCATCGCTCGCGGTGAGGGAGCGACCCCGCTCTCGCGCGATCGCTTCCGC 240
Qy 241 TCACATTCGATCTGTTCGAACCGCGGGATGTTTGTATGGCACTGCGAGCGGTGTGTGC 300
Db 241 TCGCATTCGATCGCTCGCAACCGCGGGATGTTTCGAGCGCACCGCGCGGTGTGTGC 300

Qy 301 GCCCTTGGAGCTCCCGGTGATGCTGGAGAAACGGCTACGTGAATCGGCATCCATTTCAG 360
Db 301 GCGCTCGAGGTTCGAGGGTCTCGAGAGAGCGGATATGTGAACCGTCACTCTTCGAG 360
Qy 361 TTATTCGCGATCGTGGAGAGAAAGGGCCCGCTTTCAGAGTGGCATGTTGGGCGGGCG 420
Db 361 GTCATCGCGATCGTGAAGAGAGAGGGCACCCGCTTTCAGAGCGGCATGCTGGGCGGTGCG 420
Qy 421 GCCATTGAGGGTGGTCCCGACAGGGAACCTGGACTCTTGGTTGATGAGATGAGTGTG 480
Db 421 GCGATTGCGGGGCTCGTGTCCGACCGCATCTGGACACCTTGGTGGAGAGAGACGGCTG 480
Qy 481 TCCGTTAGCGAGGGCTTACTGCTTTCGGCTTGAAGCCCGCGCAACTGCAGGCTCGAGCC 540
Db 481 ACGGTGCGGAGCGCGCCACCGCTTTCGGGCTGGNACCGGGTGGCTGCGGACCGGGCC 540
Qy 541 CGCTCCGCGCGGACCTGCGTCTTTATCGAACTACACATTTGAACAAGGACCGATCTC 600
Db 541 CGTACGAGGGATGACCTTTCGCGCTTTCATCGAGTTGCACATCGAGCGGGCGGATCTCTC 600
Qy 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACTTCCATCGTTGGGTTTCGCGCATTCGG 660
Db 601 GAGCAGGAGAAAGTGGAGATCGGCTCGTGAAGGGATCGTGGTTCGCGCTTCGCG 660
Qy 661 GTTCCGCTCAAAGGAGAGAGCGACACCGCGGCAACCCCATGCACTTCGCCCAGGAT 720
Db 661 ATCAGGTGGAGGAG 720
Qy 721 GCGTGTGTACCCGCGCTTTCATGTTGAGGAGGTCAACCGGTTTGTCAACAGAGATCGCC 780
Db 721 GCGTGTGTGCGCGGCGCTCATGTGCGAGAGATCAATCGGTTCTGTCGAGAGATCGCG 780
Qy 781 GATGACAGTGGCTACCGTTGGCCACCTCACAGTGGGCCCCCGGTGGAGGCAACAGGCTC 840
Db 781 GACGCGCGGTGGCGACCGCTCGGCGCACCTCACGCTGAGCTTTCGTTGGGCTCAACAGGTT 840
Qy 841 CCGGGGAGGTGGACTTTCACACTGAGCCTGCGTCTTCGCAATGAGAGTTCGCTCCGCGTG 900
Db 841 CCGGGGCGGTTCGAGTTTCAGCTCGATCGATCGCCCGGAGGAGTTCGATCGGCTC 900
Qy 901 CTGATTCGACCGCATCTCGGTTCATGTCGAGGTTCGCTCCAGGCGGCTTCGAGTGGCTGCC 960
Db 901 CTGATTCGACAGGATCGAGGCGATGTTGGCAGAGTTCGCGCGCGGCGGAGTTCGAGGCG 960
Qy 961 GATGTGATGAATTTTTCATCTCAGCCCGGTGAGCTGGCTCTTACCATGTTGAGACGCC 1020
Db 961 GCGGTGAACGGGTTCTTCGCGCTCAGCCCTTCGCTCTGTTCTCGGTTGGTTCGTTGATCGC 1020
Qy 1021 GTTCGGAAGCGGCTTCGCGCTTCGAGTTTCACACCGGGATATCAGAGTGGGGCGGGC 1080
Db 1021 GTTCGGAAGCGGCTTCGCGCTTCGAGTTTCACCCATCGGATATCAGAGCGGGGCGGG 1080
Qy 1081 CACGACTCGATGTTTCATCGCCAGATCACGACGTCGGAATGGTTCGTTTCCAAAGCGGT 1140
Db 1081 CACGACTCGATGTTTCATCGCCAGATCACGACGTCGGAATGGTTCGTTTCCAAAGCGGT 1140
Qy 1141 GCTGGCGGAGCGAGTTCGCGAAGATGAGCCGATTCGATGACCTTCGCAAGGAACT 1200
Db 1141 GCGGGCGAAGCCATGTGCGCGGAGGAAATGGTCCGATTCGACGATCTCGGAAAGGGAGC 1200
Qy 1201 GAGGTTGTCTCCGGGTAAATGAAGGCACTTCACCGGTAA 1239
Db 1201 GATGTGTCTTCAGTCTGAGCGGCTTCACCGGTGA 1239
RESULT 5
US-10-289-360-3
; Sequence 3, Application US/10289360
; Publication No. US20030109013A1
; GENERAL INFORMATION:
; APPLICANT: TAKENAKA, YASUHIRO
; APPLICANT: SUZUKI, SHUNICHI
; APPLICANT: ONISHI, NORIMASA

APPLICANT: YOKOZEKI, KENZO
TITLE OF INVENTION: DNA ENCODING HYDANTOINASE, DNA ENCODING N-CARBAMYL-L-AMINO ACID
TITLE OF INVENTION: HYDROLASE, RECOMBINANT DNA, TRANSFORMED CELL, METHOD OF PRODUCING
TITLE OF INVENTION: METHOD OF PRODUCING OPTICALLY ACTIVE AMINO ACID
FILE REFERENCE: 229195US00CONT
CURRENT APPLICATION NUMBER: US/10/289,360
CURRENT FILING DATE: 2002-11-07
PRIOR APPLICATION NUMBER: PCT/JP02/02173
PRIOR FILING DATE: 2002-03-08
PRIOR APPLICATION NUMBER: JP 2001-0650814
PRIOR FILING DATE: 2001-03-08
PRIOR APPLICATION NUMBER: JP 2001-298619
PRIOR FILING DATE: 2001-09-27
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 3
LENGTH: 1239
TYPE: DNA
ORGANISM: Microbacterium liquefaciens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)..(1239)
OTHER INFORMATION:
US-10-289-360-3

Query Match 59.7%; Score 739.8; DB 15; Length 1239;
Best Local Similarity 74.8%; Pred. No. 8.4e-212;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;

QY 1 GTGACCTCGAGAAAGCGGCGCGGATGAGAAAGAGATCGGGAGCTCTCCCGG 60
DB 1 GTGACCTCGAGAGCGGCGGCGGATCGCATCGAGGAGGAGCTCTGGACTCTCTCCCGC 60
QY 61 TTCTCGGAGAGGCGGCGGCTGTACCGGCTCACCTACACTACAGAGCATCGCGCGG 120
DB 61 TTCTCGGAGAGGCGGCGGCTGTACAGTCTCACGTACACTCCGAGACGCGCGCGG 120
QY 121 CGGAAACGCTCATTCGGCTATGAAAGCGCGGCTTGAGCGTTCTGTAAGACGCACTC 180
DB 121 CGAGAGGTGATCTCGCGGCGCATGCAGCGAGCGGCGTGAGCGTCCACGAGGACGCTCTC 180
QY 181 GGAACATCATCGGCGGAGTGAAGCACTGATCGGAGCTTCTCGGATCGCGGTGCT 240
DB 181 GCAACATCATCGTCTCGGCTGAGGGGAGCGACCGGCTCTGCGGCGATCGCCTTCGGC 240
QY 241 TCACACTTCGATTCTGTCGAAACGCGGCGGATGTTGATGCGCACTCGACGCGTGGGTGC 300
DB 241 TGCACCTTCGACTCGGTTCGAAACGCGGCGGATGTTGACGCGCACCGCGGCGTGGGTGC 300
QY 301 GCCCTTGAGGCTGCCCGGCTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTGAG 360
DB 301 GCGCTCGAGGCTGCGAGGCTGCTGCAGGAGCGGATATGTGAACCGTCACTCTCTCGAG 360
QY 361 TTCATCGCATCTGTGAGAGGAGGCGGCGGCTTTCAGCATGCGCATGTTGGCGCGCGG 420
DB 361 GTCATCGCATCTGTGAAAGAGGAGGCGACCGGCTTTCAGCAGCGGCGATGCTGGCGGCTCGC 420
QY 421 GCCATTGACGGGTGCTGCGCGGACAGGAACTGCACTCTTTGCTTGTATGAGGATGAGTG 480
DB 421 GCGATCGCGGGCTGTGCTCGAGCGCGGATCTGACACCCCTGTGTGACGAAAGCGGCGTG 480
QY 481 TCCGTTAGGCGGCGCTACTGCTTTCGGCTTGAAGCGCGGCGAACTGCGAGGTGCGAGCC 540
DB 481 ACGTGTGCGGAGCGCGCGCGCTTTCGGGCTGGAACCGGCTGAGCTGCGGAGCGGCGCC 540
QY 541 CGTCCGCGGCGGACCTGCGGTGCTTTTATCGAACTACATTAAGAAACGAGACCGATCTCTC 600
DB 541 CGTACGAGGAGTACCTTTCGCGCTTTCATCGAGTTGCACTATGACATCGAGCAGGCGGCGATCTC 600
QY 601 GAGCAGGACCAATAGATCGAGTTGTAACCTCCATCGTTGGCGCTTCGCGCATTTGCCG 660
DB 601 GAGCAGGAGAGGTGAGATCGGCGCTGTCGAGCGGAGTCTGCTGGTTCGCGCGCTTCCCG 660

QY 661 GTTGGCGTCAAAAGCGAGAGCGACCAACGCGGCGACAACCCCATGCACTTCCGCCAGGAT 720
DB 661 ATCACGGTGGAGGCGAGGAGCGACCAACGCGGCGAGCACCCCATGCACTTCCGCCAGGAC 720
QY 721 GCGCTGGTACCGCGGCTCTCATGGTGGAGGAGGTTCAACGGGTTCTGTCACGAGATCGCC 780
DB 721 GCGCTGGTACCGCGGCGCTCATGGTGGAGAGATCAATCGGTTCTGTCACGAGATCGCG 780
QY 781 GATGGCACAGTGGCTACCGTTGGCCACCTCACAGTGGCGGCCCGGTTGGAGGCAACCAAGGTC 840
DB 781 GACGGACCGTGGCGACCGTTCGGCCACCTCACGGTGACCCCTGGTGGGCTCAACCAAGTT 840
QY 841 CCGGGGAGGTGGAATTCACACTGGACCTGCGCTTCCGCATGAGAGATCGCTCCGCGTG 900
DB 841 CCGGGGCGGTGAGTTACGCTCGATCTCGCATCGCCACGAGGAGTCGATCCGCGCTC 900
QY 901 CTGATCGACCGCATCTCGGTCATGGTTCGGGAGGTGCGCTCCAGGCGGCTGCTGCTGCC 960
DB 901 CTGGTTCGACAGGATCGAGGCGATGGTGGCAGAGTCCCGCGCGCGGAGTCGAGGCGC 960
QY 961 GATGTGATGAATTTTTCATCTCAGCCGCGTGCACTGGCTCTTACCATGGTGGAGCGCC 1020
DB 961 GCGGTGAACGGGTTCTTCGGGCTCAGCCCTGTCTGCTCTCTCCGGTGGTGGATCGC 1020
QY 1021 GTTCCGAAAGCGGCTTCGGCTTTCAGTTTTCACACACCGGAGATATCAGCAGTGGGCGGCG 1080
DB 1021 GTGCGGAGCGGCGGCTCCGAACTTCGGCTTTCACCATCGGACATCAGAGCGGCGAGGG 1080
QY 1081 CAGGACTCATGTTTATCCCGCAGGTCACGAGCGTGGAAATGTTTTGTTTCAAGCCGT 1140
DB 1081 CAGGACTCATGTTTATCCCGCAGATCACCAGCTCGGAAATGTTTGTCTCCCGCGCCG 1140
QY 1141 GCTGGCGGAGGCGGCTTCCGAAAGATGGACCGATTCGATGACCTTCGCAAGGAACT 1200
DB 1141 GCGGGCGAAGCATGTGCGGAGGATGTCGATTCGAGATTCGAGATTCGCGAAGGGGAGC 1200
QY 1201 GAGGTTGCTCTCGGGTAATGAGGCACTTGACCGGTA 1239
DB 1201 GATGTGCTCTTTCACGTCGTCGCGGCTTGACCGGTGA 1239

RESULT 6
US-09-950-772-7
Sequence 7, Application US/09950772
Patent No. US20020102713A1
GENERAL INFORMATION:
APPLICANT: SUZUKI, Shunichi
APPLICANT: ONISHI, No. US20020102713A1Imasa
APPLICANT: YOKOZEKI, Kenzo
TITLE OF INVENTION: 5-SUBSTITUTED HYDANTOIN RACEMASE, DNA CODING FOR THE RACEMASE, A
FILE REFERENCE: 212318US0
CURRENT APPLICATION NUMBER: US/09/950,772
CURRENT FILING DATE: 2001-09-13
PRIOR APPLICATION NUMBER: JP2000-278571
PRIOR FILING DATE: 2000-09-13
PRIOR APPLICATION NUMBER: JP2001-65815
PRIOR FILING DATE: 2001-03-08
NUMBER OF SEQ ID NOS: 12
SOFTWARE: PatentIn version 3.1
SEQ ID NO 7
LENGTH: 3343
TYPE: DNA
ORGANISM: Microbacterium liquefaciens
US-09-950-772-7

Query Match 59.7%; Score 739.8; DB 9; Length 3343;
Best Local Similarity 74.8%; Pred. No. 1e-211;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;

QY 1 GTGACCTCGAGAAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 60
DB 2105 GTGACGCTGACGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2164

QY 61 TTCTCGGCAAGAGCCCGCGTGTACCCGCGTGACCTACACTCCAGAGCATCGCCCGG 120
Db 2165 TTCTCGGTGAAAGGGCCCGCGTGACACGTCTACGTACACTCCGGAGCACCCCGCGG 2224
QY 121 CGGGAACCGCTCATTTGCGGCTATGAAAGCGCGCCTTCAGAGGTTTCGTGAAGACGCATC 180
Db 2225 CGAGAGGTGATCGTCGCCCATGACGCGACGGGGCTGAGCGTCCACGAGGACGCTCTC 2284
QY 181 GGAACATCATCGCCCGACGTGAAGGCACTGATCCGGAGCTTCCTGCGATCGCGGTGCGT 240
Db 2285 GGCACATCATCGGTGCGGCTGAGGGAGCGACCCCGCTCTGCGCGCATCGCTTCGCG 2344
QY 241 TCACACTTCGATCTCTGCGAAGCGCGGATGTTGATGCGCACTGCAGCGCTGTGTGC 300
Db 2345 TCGCACTTCGATCTCGGTGCGCAACCGGGATGTTTCGACGGCACCGCGGCGTGTGTGC 2404
QY 301 GGCCTTAGGCTGCCCGGTGATGCTGGAGAACGGCTACGTGAATCGGCATCCATTTGAG 360
Db 2405 GCGCTGAGGCTGCGAGGCTGTCGAGGAGAGCGATATGTGACCGTCACTCTCTCGAG 2464
QY 361 TTCATCGCATCTGAGAGAGAAAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
Db 2465 GTCATCGCATCGTGAAGAGAGGGACCCCGCTTCAGCAGCGCATGCTGGCGCGTGC 2524
QY 421 GCATTGCAAGGTTGTTCCGCCACAGGAACTGGACTCTTTGGTTGATGAGGATGGAGTG 480
Db 2525 GCGATCGCGGGCTGTTGTCGACGCGCATCTGGACACCCCTGGTGGACGAACGCGGTG 2584
QY 481 TCCGTTAGCGAGCGGCTACTGCTTCGCGTTTGAAGCCGGGCGAACTGACGCTCGAGCC 540
Db 2585 ACGGTGCGAGCGCGCACGCGCTTCGGGCTGGAACCGGTGAGCTGGAGCGGGGCC 2644
QY 541 CGCTCGCGCGGACCTCGTGTGTTTTATCGAACTACACATTTGAACAAGGACCGATCTTC 600
Db 2645 CGTACAGGGATGACCTTCGCGCTTCATCGAGTTGCAATCGAGCAGGCGCGATCTCTC 2704
QY 601 GAGCAGGCAATAGAGATCGAGTTGTAACCTCATGTTGGCGTTTCGCGCATTTGGG 660
Db 2705 GAGCAGGAAGGTGAGATCGCGCTGCTGAGCGGGATCGTCGGTGTCCGCCCTTCGG 2764
QY 661 GTTTCGCGTCAAAAGCAGAGACCAACGCGGCAAAACCCCATGCACTGCGCCAGGAT 720
Db 2765 ATCACGTTGAGGACGAGAGACCAACGCGGACGACCCCATGCACTGCGGAGGAC 2824
QY 721 GCGTGGTACCGCGCTCTCATGTGAGGAGGTCAACCGGTTTCGTCAACGAGATCGCC 780
Db 2825 GCGCTGGTCCGCGCGCTCATGTGTGCGAGAGATCAATCGGTTTCGTCAACGAGATCGCG 2884
QY 781 GATGACACAGTGGCTACCGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACGAGTC 840
Db 2885 GACGACGAGTGGGACCGCTGGCCACCTCACGAGTACCCCTGGTGGGCTCAACAGGTT 2944
QY 841 CCGGGGAGGTGAGCTTACACTGACCTGCGTTCTTCGCATGAGAGTTCGCTCGCGCTG 900
Db 2945 CCGGGGGCGTCGAGTTACGCTCGATCTGCGATCGCCCAACGAGAGTTCGATCGGCTC 3004
QY 901 CTGATCGACCGCATCTCGGTCTATGTGCGGAGGTGCGCTCCAGCGCGGTGTGGTGCC 960
Db 3005 CTGGTGCAAGGATCGAGGCGATGTGTGGAGAAGTTCGCGCGCGCGCGAGTTCGAGGCC 3064
QY 961 GATGTGATGAATTTTCAATCTCAGCCCGGTGACGTGGCTCTTACCATGTTGACGCC 1020
Db 3065 GCGGTGAACGGGTTCTTCGCGCTCAGCCCTGTCGGTCTCTCTCCGGTGGTCTGGATGCG 3124
QY 1021 GTTCGGAAGCGGCTTCGCGCTTTCAGTTTCACACACCGGGATATCAGCAGTGGGGCGGCG 1080
Db 3125 GTGCGGACGCGCGTCCGAACTCGGCTTACCCATCGGACATCACAGAGCGGGCAGGG 3184
QY 1081 CAGGACTCGATGTTCTATGCCCAAGGTACGAGAGGTTCGGAATGGTTTTGTTCCAAAGCGGT 1140
Db 3185 CACGACTCGATGTTCTATCCGCCAGATCACCGAGTGGAAATGGTGTTCCTCCCGAGCGCG 3244

QY 1141 GCTGCGGAGCCAGCTTCCGAAAGATGACCGATTCGATGATCCTTCGAAAGAACT 1200
Db 3245 GCCGGCGAAGCCATGTGTCGCGAGGAATGGTCCGATTCGACGATCTCGGAAGGGAGC 3304
QY 1201 GAGGTTGTCTCCCGGTAAATGAAGCACTTGACCGGTAA 1239
Db 3305 GATGTGTCTCTACGTCGTGACGCGCTTGACCGGTGA 3343
RESULT 7
US-10-289-360-5
; Sequence 5, Application US/10289360
; Publication No. US20030109013A1
; GENERAL INFORMATION:
; APPLICANT: TAKENAKA, YASUHIRO
; APPLICANT: SUZUKI, SHUNICHI
; APPLICANT: ONISHI, NORIMASA
; APPLICANT: YOKOZEKI, KENZO
; TITLE OF INVENTION: DNA ENCODING HYDANTOINASE, DNA ENCODING N-CARBAMYL-L-AMINO ACID
; TITLE OF INVENTION: HYDROLASE, RECOMBINANT DNA, TRANSFORMED CELL, METHOD OF PRODUCING
; TITLE OF INVENTION: METHOD OF PRODUCING OPTICALLY ACTIVE AMINO ACID
; FILE REFERENCE: 229195USOCONT
; CURRENT APPLICATION NUMBER: US/10/289,360
; PRIOR FILING DATE: 2002-11-07
; PRIOR APPLICATION NUMBER: PCT/JP02/02173
; PRIOR FILING DATE: 2002-03-08
; PRIOR APPLICATION NUMBER: JP 2001-0650814
; PRIOR FILING DATE: 2001-03-08
; PRIOR APPLICATION NUMBER: JP 2001-298619
; PRIOR FILING DATE: 2001-09-27
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 3343
; TYPE: DNA
; ORGANISM: Microbacterium liquefaciens
US-10-289-360-5
Query Match 59.7%; Score 739.8; DB 15; Length 3343;
Best Local Similarity 74.8%; Pred. No. 1e-211;
Matches 927; Conservative 0; Mismatches 312; Indels 0; Gaps 0;
QY 1 GTGACCTTCAGAAAGCGCAAGCGCGCATTTGAGAAAGAGATCCGGGAGCTCTCCCGG 60
Db 2105 GTGACGTGTCAGAGCGCGCGCGCATCGCATCGAGGAGAGCTCTGACTCTCTCCCGC 2164
QY 61 TTCTCGGAGAAAGCGCCCGGTGTTACCGGCTGACCTTACACTCCAGAGCATGCGCCCGG 120
Db 2165 TTCTCGGTGGAAGGGCCCGGTGACACGTCTACGTACACTCCGGAGCAAGCGCCCGG 2224
QY 121 CGGGAACGCTCATTTGCGGCTATGAAAGCGCGCCCTTGAGCGTTTCGTGAAGACGCATC 180
Db 2225 CGAGAGGTGATCGTCGCGCCCATGACGCGAGCGGGCTGAGCGCTCCAGAGACGCTCTC 2284
QY 181 GGAACATCATCGCGCGGACGTGAAGGCACTGATCCGAGAGCTTCTTCGCATCGCGGTGCGT 240
Db 2285 GGCACATCATCGTTCGCGGTGAGGGAGCGACCCCGCTCTGCGCGCATCGCTTCGCG 2344
QY 241 TCACACTTCGATCTGTCCGAAAACGCGCGGATGTTTGTATGGACATGACAGGCGTGTGTGC 300
Db 2345 TCGCACTTCGACTCGGTCCGCAACGCGCGGATGTTTCGACGGCACCGCGGCGGTGTGTGC 2404
QY 301 GCCCTTAGGCTGCCCGGTGATGCTGGAGAACCGCTACGTGAATCGGCATCCATTTGAG 360
Db 2405 GCGCTGAGGCTGCGAGGCTGTCGAGGAGAGCGATATGTGACCGTCACTCTCTCGAG 2464
QY 361 TTCATCGCATCTGAGGAGAAAGGGCCCGCTTCAGCAGTGGCATGTTGGCGCGCGG 420
Db 2465 GTCATCGCATCGTGAAGAGAGGGACCCCGCTTCAGCAGCGCATGCTGGCGCGTGC 2524
QY 421 GCATTGCAAGGTTGTTCCGCCACAGGAACTGGACTCTTTGGTTGATGAGGATGGAGTG 480
Db 2525 GCGATCGCGGGCTGTTGTCGACGCGCATCTGGACACCCCTGGTGGACGAACGCGGTG 2584

```
QY 481 TCGTTAGGAGCGGCTACTGCTTCGGTTGAAGCCGGCGAATGCGAGGTGCGACCC 540
Db 2585 ACGGTGCGGAGCGGCGACGGCTTCGGGTGGAACCGGGTGAAGTTCGGAGCGGCGCC 2644
QY 541 CGCTCCGCGGCGGCGCTGCTGCTTTTATCGAACTACACATTTGAACAGGACCGATCCTC 600
Db 2645 CGTACGAGGATGACCTTCGGGCTTCATCGAGTTGCACATCGAGAGGGCGGATCCTC 2704
QY 601 GAGCAGGAGCAATAGAGATCGGAGTTGTAACTCTCCATCGTTGGCGTTCCGCAATGCGG 660
Db 2705 GAGCAGGAGAGTGGAGATCGGCTGTCGACGGGATCGTGGTTCGCGGCTTCGG 2764
QY 661 GTTCCGCTAAAGCAGAGGACCGCGCGGACAAACCCCAATGCACTTGGCGCAGGAT 720
Db 2765 ATCAGGTGAGGGCAGAGGAGCAGCAGCGCGGAGCAGACCCCATGCACTTGGCGCAGGAC 2824
QY 721 GCCTGTATCCCGCGCTCTCATGTTGAGGAGGTCAACCGGTTCTGTCACAGGATGCGC 780
Db 2825 GCCTGTGTCGGCGGCTCATGTTGCGAGAGATCAATCGGTTCTGTCACAGGATGCGG 2884
QY 781 GATGGCAGAGTGTACTCGTTGGCCACTCACAGTGGCCCCCGTGGAGGCAACAGGTC 840
Db 2885 GACGGCAGGTGGGACCGTTCGGCCACTCACGGTGAACCCCTGGTGGGCTCAACAGGTT 2944
QY 841 CCGGGGAGGTGGAATTCACACTGGACCTCGGTTCTCCGATGAGGAGTGCCTCCGCGTG 900
Db 2945 CCGGGGCGCTCGAGTTCACTCGCTCGATCTCGGATCGCCCCACAGGAGTGCATCCGCGTC 3004
QY 901 CTGATCAGCCGATCTCGGTGATGTTGTCGGGAGGTGCGCTCCAGCGCGGTGCGTGCC 960
Db 3005 CTGGTGCACAGGATCAGAGCGGATGGTGGCAGAGTGGCGCGCGCGGAGTCGAGGCC 3064
QY 961 GATGTGATGAATTTTCAATCTCAGCCCGGTGCGCTGCTCTACCATGTTGGAGGCC 1020
Db 3065 GCGGTGAACGGGTTCTTCGGCTCAGCCCTGTCGGTCTGCTCGGTGCTGATGCGC 3124
QY 1021 GTTCCGGAAGCGCTCGGCTTCGAGTTTCAACACCGGATATCAGAGTGGGGCGGC 1080
Db 3125 GTGCGGAGCGCGCTCGGAATCTCGGCTTCAACCATCGGACATCAGAGCGGGCAGGG 3184
QY 1081 CAGGACTCGATGTTTCATCGCCGAGTCAAGGATGCGGATGCTTTCGATCAAGCCGT 1140
Db 3185 CAGGACTCGATGTTTCATCGCCGAGTCAAGGATGCGGATGCTTTCGATCAAGCCGT 3244
QY 1141 GCTGGCGGAGCAGCTTCCGGAAGATGACCGATTTTCGATGACCTTCGCAAGGAACT 1200
Db 3245 GCGGGGGAAGCCATGTGCGGAGGATGTCGATTTTCGATGTCGGAAGGAGCG 3304
QY 1201 GAGGTTGCTCTCCGGTAATGAAGGCACTTTGACCGGTAA 1239
Db 3305 GATGTGCTCTTCACTCGTGCAGCGGCTTTGACCGGTGA 3343
```

RESULT 8

```
US-10-398-221-10
; Sequence 10, Application US/10398221
; Publication No. US200400185141
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 10
; LENGTH: 1163020
```

```
; TYPE: DNA
; ORGANISM: Listeria innocua
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a, g, c or t/u
US-10-398-221-10

Query Match      17.8%; Score 220.6; DB 16; Length 1163020;
Best Local Similarity 50.2%; Pred. No. 3.4e-55;
Matches 572; Conservative 0; Mismatches 564; Indels 3; Gaps 1;

QY 73 GGCCTCCGGTGTATACCGGCTGACCTACATCCAGAGATGCCCGCGCGGGAACCGCTC 132
Db 478591 GGTTCAGGGAACAACTCGGCTTACATATAGCAAAAGAACCTCGCGCGCTAATTTATTA 478650
QY 133 ATTTCGGCTATGAAGCGGCGCTTTCGAGGTTTCGTGAGACGACATCCGGAACATCATC 192
Db 478651 AAGAAGAAATGGCTAAAGTAGCCCTTACTGTCTCAGAAGATCGGATTGGAATATCTAC 478710
QY 193 GGCCTGAGTGAAGGCACTGATCCGAGCTTCTTCGATCCGCTTCGCTTCACACTTTCGAT 252
Db 478711 GGAAGACTAGAGGCGACAAATCCAGACATACCAAGAGTAATCTCGGTTCTCATTTTGC 478770
QY 253 TCTGTCCGAAACCGCGGATGTTTGTATGACATTCGAGGCGGTGTGTGCGCCCTTGAGGCT 312
Db 478771 TCTGTCCAAATGCTGCTTTTACGCTTCCAGAGGTGTTATTAACCTGGGCTTGAAGTA 478830
QY 313 GCCCGGCTGATGCTGAGAACGCGCTAGTGAATCGGCAATCCATTTGAGTTTCATCGCGATC 372
Db 478831 GCTAGCGTTTTCATGAACAGCAAAATAAAACCCCATTTTCCACTTGAAATCATCGCAATG 478890
QY 373 GTGAGGAGGAAGGGGCCCGCTTCAGCAGTGGCATGTGGGCGCGCGGCAATTCGAGGG 432
Db 478891 GTGGAAGAGAGGGAATCTCGTTTCGAGCGGGAATCTACTTGCCTCCGGAACAATACAGGA 478950
QY 433 TTGTCGCGCAGAGGAACTGGAATCTTTGTTGATGAGAGATGAGTGTCCGTTAGGCGAG 492
Db 478951 AAAGTTTACAAAGAAATGCTACATGAATATGAAGATATAAATGTTGTTACTGCGCGAGAA 479010
QY 493 GCGGCTACTGCTTCGCTTGAAGCGGCGGGAATTCAGGCTGCAGGCTGCAGCCGCTTCGCGCG 552
Db 479011 GCTATGGCAACCTAGGTTTGAACGAAATCAAGTTTCATACAGCCATTCGCTCCAAAGAT 479070
QY 553 GACCTCGCTTTTATCGAACTTACATTTGAACAGGACCGGATCTCTCGAGCAGGAGCAA 612
Db 479071 TCGCTCAAAGCTTTCATTTGAACTACATCGAACAAGGCCAGTCTCTAGAAATGCAAT 479130
QY 613 ATAGAGATCGGAGTTGTAACCTCCATCTGTTGGCGTTTCGCGCAATTCGCGGTTGCGCTCAA 672
Db 479131 GAAAGATGTTGCGCTAGTTCGACACGCTGTTGGTTTAAAGAAATAAAGTTACCGTAAAG 479190
QY 673 GGCAGAGGCGACCGCGCGGCAACCCCATGACCTCGCGCAGGATGGGCTGGTATCCC 732
Db 479191 GGTCAAAGCAGGTCACCGCGGAACTACCCCAATGCTAGAAACGAAAGATGCTCTAAGTGCA 479250
QY 733 GCGGCTCTCATGTTGAGGAGGTTCAACCGGTTCTGTCACAGAGATCCCGATGCGACAGTG 792
Db 479251 GCTGTCCAAATTTTAAAGCTTCTGAACTAGCTATCCAGAAAGCGCGGAACTGTC 479310
QY 793 GCTACCGTTGGCCACCTCAGTGGCCCCCGGTGGAGGCAACCAAGTCCCGGGGAGGTG 852
Db 479311 TTAACAAATTTGGCAAACTTAATGTTACTACCGAAGCGGCTTAATGTAATACCAACAAGTC 479370
QY 853 GACTTCACACTGACCTGCTGTTCTCGGATGAGGAGTGCCTCGCGGTGCTGATCGACCGC 912
Db 479371 GTTTTCACAGTAGATATTCGAGCAAAAGACGAAATCCACGTTCCAAAATACATTTAGAAAA 479430
QY 913 ATCTCGGTCATGTCGCGGAGGTGCGCTCCAGGCGGCTGTGCTGCCGATGTTGGATGAA 972
Db 479431 A-----CAAAAAGATTTATCAAGCTTTTGAAAAAAGCGGTTATTTATGTTGAAATAGAGAT 479487
QY 973 TTTTTCATCTCAGCGCGGTGCGAGTGGCTCTCTACCATGTTGAGACGCGCTTCGCGAAGCG 1032
```

```
Db 479488 ATGCTCTATGAATAAACACACATTTATCAACAGAAATTCATCAGCATTTGACCGAAGC 479547
Qy 1033 GCCTCGCGCTTCAGATTACACACCGGGATATCAGCAGTGGCGGGGCCACGACTCGATG 1092
Db 479548 GCCGATAAACTTGGTCTTAAATATCGAACAAATGGTTAGCGGCGCAGGACACGATGCAATG 479607
Qy 1093 TTCATCGCCCGCAGGTACGAGACGTCCGAATGGTTTCGTTCCAAAGCGGTGCTGCGCGGAGC 1152
Db 479608 ATTTTCGCAAGTTTAAACCGAAGTAGGACTATTTTGTATACCAAGCCACAAGGTTATAAGC 479667
Qy 1153 CACGTTTCCGGAAGATGGACCATTTTCGATGACCTTCGCAAGGAACTGAGGTGTGTCCT 1211
Db 479668 CATGCACCAAGAAGATGGACCGATTACGACAAAGCTCCAAAAAGGCATCGAAGTCGTACT 479726

RESULT 9
US-10-398-221-2058
; Sequence 2058, Application US/10398221
; Publication No. US200400185141
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 2058
; TYPE: DNA
; ORGANISM: Listeria innocua
US-10-398-221-2058

Query Match 17.8%; Score 220.6; DB 16; Length 3011208;
Best Local Similarity 50.2%; Pred. No. 4.2e-55;
Matches 572; Conservative 0; Mismatches 564; Indels 3; Gaps 1;

Qy 73 GCGCCGGTGTACCGCGTGACCTACACTCCAGACATGCCGCGCGGGGAAACGCTC 132
Db 568857 GGTCAAGGAAACAACTCGGCTTACATATAGCAAGAAAGACCTCGGCGCGGTAAATTATTTA 568916

Qy 133 ATTGCGGCTATGAAGCGCGCGCTTGGAGGTTCTGTGAAGACGCACTCGGAAACATCATC 192
Db 568917 AANGAAGAAATGGCTAAAGTAGGCCTTACTGTCTCAGAGATGCGATTGGAATATCTAC 568976

Qy 193 GCGCGACGTGAAGGCACTGATCGGAGCTTCCTGCGATCGCGTGGTTTCACTTCGAT 252
Db 568977 GGAAGCTAAGAGGCGCAATCCAGACATACCAGCAGTAATCGTCGTTCTCATTTGAC 569036

Qy 253 TCTGTCCGAAACGGCGGATGTTTGTATGCACTGCAAGGCGTGGTGGCCCTTGGGCT 312
Db 569037 TCTGTGCCAAATGGTGGTGCTTTTGACGCTCCAGCAGGTGTTATTAACCTGGGCTGAAGTA 569096

Qy 313 GCGCGGCTGATGCTGAGAACGGCTACGTGAATCGGCATCCATTTGAGTTTCATCGGATC 372
Db 569097 GCTAGGCTTTTTCATGAACAGCAATAAACCCATTTTCCACTTGAATCATCGCATG 569156

Qy 373 GTGAGGAGAAAGGGCCCGCTTCAGCAGTGGCAATGTTGGGCGCGCGGCCCAATGCGAGG 432
Db 569157 GTGGAAGAAGAGGATCTCGTTTCGGAGCGGGAATCTTTCCTCCCGCAAAATTAACAG 569216

Qy 433 TTGTTGCCGACAGGAACTGGACTCTTTGGTTGATGAGATGAGTGTCCGTAGGCAG 492
Db 569217 AAAGTTACAAAAGAAATGCTACATGAATGAAGATATAAATGGTGTACTGCGCGAGAA 569276

Qy 493 GCGGCTACTGCTTCCGCTTGAAGCGCGGCAACTGCAGGCTGCAGCCGCTCCGCGCG 552
```

```
Db 569277 GCTATGGCAACCTAGGATTTTACGCAAAATCAAGTTTCATACAGCCATTCGGTCCAAAGAT 569336
Qy 553 GACCTGCGTGTCTTTATCGAACTACACATTTGAAACAGGACCGATCCTCGAGCAGGAGCAA 612
Db 569337 TCCGTTAAAGCTTTTCAATTTGAATACATCAGCAAGAGCCCGAGTCTTAGAAAATCGAAT 569396
Qy 613 ATAGAGATCGGAGTTGTAACTCCATCGTTGGCGGTTCCGCGATTCGCGGTTGCGGTCCTCAA 672
Db 569397 GAAGATGTGGCTAGTTGACACGGTCTGTTGGTTTAAACAGAAATAAAAGTTTACCGTAAAG 569456
Qy 673 GCGAAGACGACACCGCGGCAACACCCCATGACCTTCGCGCAGGATGCGGTGCTGATGCC 732
Db 569457 GGTCAAGCAGGTACGCGCGGAATCTACCCCAATGCTAGAACGAAAAGATGCTCTAAGTGCA 569516
Qy 733 GCGGCTCTCATGTTGAGGAGGTCAACCGGTTCTGTCACAGGATCGCGATGGCCAGTG 792
Db 569517 GCTGTCCAAATTTCTTAATAGCTTCTTGAACCTAGCTATCCAAAGAGGCGGGAATGTC 569576
Qy 793 GCTACCGTTGGCCACCTCACAGTGGCCCCCGGTGGAGGCAACCAAGGTCCCGGGGAGGTG 852
Db 569577 TTACAAATGGCAACTTAATGTCTACCGGAACGGGCTAATGTATATACCAACAAAGTC 569636
Qy 853 GACTTCACACTGGACCTCGGTTCTCCGATGAGGAGTCGCTCCGCGTGTGATCGACCGC 912
Db 569637 GTTTTCACAGTAGATATTCGAGCAAAAGACGAAATCCACGTCCAAAATATACATTAGAAAA 569696
Qy 913 ATCTCGTCTATGTTGCGGAGGTGCGCTCCAGGCGGTGTGGCTGCGCATGTGATGAA 972
Db 569697 A---CAAAAAAAGTTATTTCAAGCTTTTCAAAAAAACGGTATTTATGTGGAATATAGAAAT 569753
Qy 973 TTTTTCATCTCAGCCCGGTGCAGTGGCTCTCTACCATGGTGAGCGCGTTTCGCGAAGCG 1032
Db 569754 ATGCTCTATGAATAACCAACACATTTATCAACAGAAATTCATCAAGCATTTGACCGAAGC 569813
Qy 1033 GCCTCGGCTTTCAGTTTCACACACCGGATATCAGCAGTGGGGCGGCGCACGACTCGATG 1092
Db 569814 GCGGATAAACTTGGTCTTTAAATATCGAAACATGTTAGCGGCGCAGGACACGATGCAATG 569873
Qy 1093 TTCATCGCCCGAGTCAAGGATCGGAATGGTTTTCGTTCCAAAGCGGTGCTGCGCGGAGC 1152
Db 569874 ATTTTCGCAAGTTTAAACCGAAGTAGGACTTATTTTGTACCAAGCCCAAAAGGTATAAGC 569933
Qy 1153 CACGTTCCCGAAGATGACCGGATTCGATGACCTTCGCAAGGAACTGAGGTTGTCTCT 1211
Db 569934 CATGCACCAAGAAATGACCGGATTACGACAAAGCTCCAAAAGGCATCGAAGTCGTACT 569992

RESULT 10
US-10-398-221-3817/c
; Sequence 3817, Application US/10398221
; Publication No. US200400185141
; GENERAL INFORMATION:
; APPLICANT: KUNST, Frederik
; APPLICANT: GLASER, Philippe
; TITLE OF INVENTION: Listeria innocua, genome and applications
; FILE REFERENCE: 344 702 - US
; CURRENT APPLICATION NUMBER: US/10/398,221
; CURRENT FILING DATE: 2003-03-27
; PRIOR APPLICATION NUMBER: PCT/FR 01/03 061
; PRIOR FILING DATE: 2001-10-04
; PRIOR APPLICATION NUMBER: FR 00/12 697
; PRIOR FILING DATE: 2000-10-04
; NUMBER OF SEQ ID NOS: 4025
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 3817
; LENGTH: 3880
; TYPE: DNA
; ORGANISM: Listeria monocytogenes 4b
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(end)
; OTHER INFORMATION: n can be any nucleotide: a,g,c or t/u
US-10-398-221-3817
```

Query Match 12.8%; Score 158.6; DB 16; Length 3880;
Best Local Similarity 49.8%; Pred. No. 4e-37;
Matches 429; Conservative 0; Mismatches 429; Indels 3; Gaps 1;

QY 351 TCCATTGAGTTCATCCGATCTGGAGGAGGAGGCGCGCTTCAGCAGTGGCATGTT 410
Db 3855 TCCGCTTGAATTAATTCGATGGTGGAGAGAGGCGCTCGTTTGGCGCTGGCCCTCT 3796
QY 411 GGGCGGCGCGGCATTCAGGGTGGTCGCCGACAGGAACTGGACTCTTTGGTTGATGA 470
Db 3795 TGCTTCAGTACATTAATCTGGCAAGTTACAAAGAAATGCTCATGAATGAAGATAT 3736
QY 471 GGATGGAGTTCGGTTAGGAGCGGCTACTGCTTCGGCTTGAAGCGCGGCAACTGCA 530
Db 3735 AGATGGAAATACAGCGCAGAACCAATGGCAAGTAGGATTTGATCAAAATCAAGTAGT 3676
QY 531 GGTGAGCGCGCTCCGCGGCGGACCTCGCTGCTTTATCGAACTACATTAACAGG 590
Db 3675 TACAGCTATCCGACGAAAGAAATCCGTTAAAGCATTTTATCGAACTACATTAAGCAAG 3616
QY 591 ACCGATCTCTGAGCAGGAGCAATAGAGATCGGATTTGTAACCTTCACTCGTTGGCGTTCG 650
Db 3615 ACCAGTCTAGAAAATGCCAATGAAGATGGCGCTAGTAGATACAGTAGTTCGTTTAA 3556
QY 651 CGCATTCGCGGTTGCGTCAAGGAGAGAGCGACCGCGGACCAACCCCATGCACTT 710
Db 3555 AGAAATAAAGTAACAGTAAAGGACAAAGCGCGCATGCGAGGACCAACCCCATGCTTGA 3496
QY 711 GCGCCAGGATGCGCTGTACCGCGCTCTCATGTTGAGGAGGTCAACCGTTTCGTCAA 770
Db 3495 CCGAAAGATGCGCTGTATCACTGCTGCAATTTTAGGTCAACTACAGAACTAGCTAT 3436
QY 771 CGAGATCGCGGATGGACAGTGGCTACCGTTGGCACTTCACTGAGTGGCGCGCGTGGAGG 830
Db 3435 CCAAGAGGCGCGGAACTGTATTAACAGTGGGCAAACTCAACGTCTATCCAAATGGCGC 3376
QY 831 CAACCAAGTCCCGGGGAGGTGGACTTCACTGGACCTCGCTTCTCCGATGAGGAGTC 890
Db 3375 AAACGTTATTCAGATAAAGTCTGTTTACCGTAGATATCCGAGCAAAAGACGAATTC 3316
QY 891 GCTCCGGTCTGATCGACCGCATCTCGGTTCATGTTGGGAGGTGCGCTCCAGGCGCG 950
Db 3315 CGTCCAAAATACATTAAGCAAAACAAAGAAATAT---CAAATCCGACAGAGAAACCG 3259
QY 951 TGTGGCTGCGATGTGATGAAATTTTCAATCTCAGCCGCTGAGTGGCTCTCTACCAT 1010
Db 3258 CATCACTTGGCAATAGAGATATGCTATCGAACCGCCGACCATTTTATCAAAAGAT 3199
QY 1011 GGTGAGCGCGTTTCGCAAGCGGCTCGGCTTTCAGTTTCAACACCGGATATCAGCAG 1070
Db 3198 TCATCAAGCATTAAGTGAAGTCCGACCACTCGGCTTAAATACCGAATGTTAG 3139
QY 1071 TGGGGCGGCGCAGCTCGATGTTTCATGCGCCAGGTCAAGGAGTGGAAATGTTTCTGT 1130
Db 3138 CGGGGCGGACAGATGCAATGATTTTTCAGGTTTAAACCGAAAGTGGCTTGAATTTTGT 3079
QY 1131 TCCAACCGTGTGGCGGAGCCAGCTTCCGAGAAATGAGCCGATTTGATGACCTTCG 1190
Db 3078 CCTACCAATGTTATAGCCATGCGCCGAGAAATGAGCGATATACGAAACTTCCA 3019
QY 1191 CAAAGGAATGAGTTGTCT 1211
Db 3018 AAAAGGAATCGAAGTCTACT 2998

RESULT 11
US-09-974-300-308
; Sequence 308, Application US/09974300
; Patent No. US2002014672A1
; GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; APPLICANT: Clausen, Ib Groth

; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; TITLE OF INVENTION: Expression
; FILE REFERENCE: 10085.500-US
; CURRENT APPLICATION NUMBER: US/09/974,300
; CURRENT FILING DATE: 2001-10-05
; PRIOR APPLICATION NUMBER: 09/680,598
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 60/279,526
; PRIOR FILING DATE: 2001-03-27
; NUMBER OF SEQ ID NOS: 8481
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 308
; LENGTH: 1218
; TYPE: DNA
; ORGANISM: Bacillus licheniformis
US-09-974-300-308

Query Match 12.1%; Score 150.2; DB 9; Length 1218;
Best Local Similarity 47.3%; Pred. No. 1e-34;
Matches 492; Conservative 0; Mismatches 543; Indels 6; Gaps 1;

QY 180 CGAAACATCATCGCGCGAGCTGAAGCACTGATCGGAGCTTCTCGCATCGGCTCGG 239
Db 167 CGAAATCTGTTGGAGAAATAGAGTACAGAAACCGTACGAAACCGATCTTGACCG 226
QY 240 TTCACATCTCGATTCTGTCGAAACCGCGGATGTTTGTATGGCACTGCGAGCGTGTGTG 299
Db 227 TTCACATCTGATACGTCATCAACGAGGAAATTCGAGGAGCGTACGGGATTTAGC 286
QY 300 CGCCCTTGAAGCTGCGCGGCTGATGTGGAGAACGGCTAGTGAATCGGCAATCCATTGA 359
Db 287 GAGCCTTTTGTCTGCAAAACACCTTCTGGACAGTACGGCAGCCGCAAAACGTCNAATCG 346
QY 360 GTTCATCGCATCGTGGAGGAGGAGGCGCGCTTTCAGCAGTGGCATGTTGGGCGCGC 419
Db 347 AGTCGTGTCGTGCGAGGAGGAGGAGGCGCTTTCCTTTGACTTTTGGGGATCGCG 406
QY 420 GGCATTTGACAGGCTTGGTCGCCAGAGGAACTGGACTCTTTGTTGTATGAGGATGAGT 479
Db 407 GAATTTAAACGGCTTGTGACTGTCAAAAGACCGCGCTGTGCACGATCGGAGCGGAGT 466
QY 480 GTCCGTTAGGACGAGCGGCTACTGCTTCGCTTGAAGCCGCGGCAACTGCGAGCTGCAGC 539
Db 467 ATCGATTTGACGAGCGGATGAAAGATTCGCTTT-----CGGAAACCGCGGTACAGCG 520
QY 540 CGCTCCGCGGCGGAGCTCGGCTGCTTTTATCGAACTACACATTTGAACAGGACCGATCCT 599
Db 521 CCGCTTCAGGCGGTGATGTAAGTGTCTTATCGAGCTTCATATTTGAACAGGGAGCATTTCT 580
QY 600 CGAGCAGGAGCAAAATAGAGATCGGAGTGTGTAACCTCCATCGTTGGCGTTTCGCGCATTCG 659
Db 581 TGAGGAGAGCGGCGCTCAATCGGCATTTGTACGGACATGCTCGGACAAAGACGGTTTAC 640
QY 660 GGTGCGCTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 719
Db 641 GGTATGTTCAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 700
QY 720 TCGGCTGTGTAACCGCGCTCTCATGTTGAGGAGGAGTCAACCGGTTCTGTCACAGATCGC 779
Db 701 TCGGCTGTGCGCTCTCGGCATTTGTCATTTTTCATGACGAAACAGGCTTAAGCTGCTTA 760
QY 780 CGATGACAGTGGCTTACCGTTTGGCCACCTCAAGTGGCGCGCGGTGGAGGCAACAGGT 839
Db 761 TCGGCTTTTAAACAGCAACCGTGGGAGGATTTGAAGCAAGCGAATGTCCGCAATGTCT 820
QY 840 CCGGGGGAGGTGGACTTTCACACTGACCTGCTTCTCCGATGAGGAGTTCGCTCCGCGT 899
Db 821 ATCAGGAGAGGCTGTTTTTCTTCTTCTGATCTCCGCAATCATGATGCGGTGTTGACCG 880
QY 900 GCTGATCGACCGCATCTCGTTCATGTTGCGGAGGAGTTCCTCCAGGCGGCTGCTGCTGC 959
Db 881 GTATTGCGACGATATTTTGTCTTATTTTCAGAGCTGGCGCGCAATGGAACGTTTCAGAT 940


```
RESULT 13
US-10-329-960-1
; Sequence 1, Application US/10329960
; Publication No. US2003009277A1
; GENERAL INFORMATION:
; APPLICANT: Fleischmann et al.
; TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus influenzae Rd Genome, Frag
; FILE REFERENCE: PB186P1
; CURRENT APPLICATION NUMBER: US/10/329,960
; PRIOR FILING DATE: 2003-01-02
; PRIOR APPLICATION NUMBER: US 09/643,990
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: US 08/487,429
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/426,787
; PRIOR FILING DATE: 1995-04-21
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 1830121
; TYPE: DNA
; ORGANISM: Haemophilus influenzae
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (4747)..(4747)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (9921)..(9921)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (10150)..(10150)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (29298)..(29298)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36543)..(36543)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36551)..(36551)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (36636)..(36636)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (40810)..(40810)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44416)..(44416)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44905)..(44905)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (44975)..(44975)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45593)..(45593)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (45732)..(45732)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (47036)..(47036)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51334)..(51334)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51602)..(51602)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51786)..(51786)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (51805)..(51805)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (55369)..(55369)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65309)..(65309)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (65313)..(65313)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (80024)..(80024)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (100091)..(100091)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (102696)..(102696)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (105121)..(105121)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (107248)..(107248)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (117136)..(117136)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119750)..(119750)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (119924)..(119924)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (120038)..(120038)
; OTHER INFORMATION: n equals a, t, g or c
; FEATURE:
; NAME/KEY: misc_feature
```

```
/ LOCATION: (121344)..(121344)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (122167)..(122167)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145058)..(145058)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc feature
/ LOCATION: (152530)..(152530)
```

```
Query Match 9.8%; Score 121.2; DB 15; Length 1830121;
Best Local Similarity 43.8%; Pred. No. 2.7e-25;
Matches 525; Conservative 0; Mismatches 673; Indels 0; Gaps 0;

QY 28 CGCATTTGAGAAAGAGATCGGGAGCTCTCCCGTTCCTCGGCAGAGGCCCGCGTGTACC 87
DB 608660 CGTGTTCAAATTTAAATGAAAACCTGGCTTTTATTCATCATGATCCAAACGAGCTGACT 608719

QY 88 CGGCTGACCTACACTCCAGAGCATGCCGCCGCGGGAACCGCTCATTTGCGGCTATGAAA 147
DB 608720 CGCTTAGCTTTTACAGAGAGGATGAAAAGGCCCAATAATATGATTTATGAAATTAAGTAA 608779

QY 148 GCGGCGGCTTGAGGCTTGTGAAGACGCACTCGGAAACATCATCGGCCGAGCTGAAGGC 207
DB 608780 GAATATGATTTTGTCTATTTCGTCGAGATTCAATTGGAAATCTTTTATTTCGTAAGGCAGGT 608839
```

```
QY 208 ACTGATCCGAGCTTCCTCGCATCGCGTTCGTTCCACACTTCGATTCGATTCGAAACGCGC 267
DB 608840 AAAGAGAGATTTTACCTGCGAGTTGCAATTCGGAATCACATATATGATCTGTTGTGATGCT 608899

QY 268 GGGATGTTTGTATGCACTGCAAGCGTGTGTGCGCCCTTGAGGCTGCCGCGGTGATGCTG 327
DB 608900 GGTAAATTTGATGTCCTTTGCGGATCTGTTGCTGGTTTGGAAATTTCTTTTGGCAATTTGTG 608959

QY 328 GAGAACGGCTAGGTGAATCGGCATCCATTTGAGTTTCATCGCATCGTGAGGAGGAAGGG 387
DB 608960 GAACAGAAATATTCAGACTCGTTATCCTTTAGAAATTTGATTAATTTTACTTGTGAAGAACTG 609019

QY 388 GCCGCTTTCAGCAGTGGCATGTTGGCGCGCGGGCCATTGCGAGGTTGCTGCGCCGACAGG 447
DB 609020 AGTAGATTTAAATTTGCTACATTTGGGTAGTAAAGTTATGTGGCATAGTAAATCAAGAA 609079

QY 448 GAACTGGACTCTTTGGTTGATGAGATGGAGTGTCCGTTAGGCAGCGCGCTACTGCTCTTC 507
DB 609080 AAATTTAAGTTTCATTACGTGATAAACAAGAAAGGGTTTATCAGAAGCTATGCTGGAAGTA 609139

QY 508 GCGTTGAAGCCGGGGAACTGCAGCTGCAGCCCGCTCCGCGCGGACCTCGGTGCTTTT 567
DB 609140 GGAATGAAATTTTAAATTTGTTTAAATCAAGCAAAACGTGATGCAAAAGGAATTTTAAATGTTT 609199

QY 568 ATCGAACTACACATTTGAAACAAGGACCGATCCTCGAGCAGGAGCAAAATAGAGATCGAGTT 627
DB 609200 TTTGAACTTCATATAGAACAAAGGCCCTCGTTTAGAGATGAGGAAACAAATAGGTGTT 609259

QY 628 GTAACTCTCCATCGTTGGCGTTTCGCGCATTTGCGGTTGCGGTCAAGAGGAGGAGCAAC 687
DB 609260 GTGACAGGTAATGCTGCCCGCATTTCTGTCGAATTTGTTAAATTTAAAGGACAAAGCGATCAT 609319

QY 688 GCCGCACAACCCCATGACCTCGGCCAGAGATGGCTGGTACCGCGCGCTCTCATGTGTG 747
DB 609320 TCAGGAGCAACAGCAATGTCATTTATCGTCATGATGCATTTGTTAGGAGGATCTGAGTTATCA 609379

QY 748 AGGAGGTCACACCGGTTCTGTCACAGATGCCCATGSCACAGTGGCTACCGTTGGCCAC 807
DB 609380 CTTGCTATTGAGCAGCTGCTATTCAAGCTGGACATTTCTACAGTAGCTACAGTAGGTAAAT 609439

QY 808 CTCAGTGGCCCCCGGTGGAGGCAACACAGGTCCCGGGGAGGTGGACTTCACACTGGAC 867
DB 609440 ATTACAGCTAAACCAGGAGTAATGAATGTTGTGCCAGGATATTGCGAATTTATTAGTAGAT 609499

QY 868 CTGCTTCTCCGATGAGGAGTCCCTCGCGGTGCTGATCGACCGCATCTCGTCTCATGTGC 927
DB 609500 ATTCGAGGTACACATGTACAGCTAGAGATTTCTGTATTTTGAATTTATTACAAGAGGAAT 609559

QY 928 GGCAGGTCGCTCCCGCGCGGTGTGCTGCCGATGTGGATGAATTTTCAATCTCAGC 987
DB 609560 AGTAAAGTTTCAGAAAAAAGAGGATTTAATTAAGTTACAACTATTTCAAAAGATAAAT 609619

QY 988 CCGGTGCAGCTGGCTCTCTACCATGGTGGACCGCTTCGCGAAGCGGCTCGGCTTCGAG 1047
DB 609620 CCAATAATATTACCTGAAATATGTAACCAAAATAGTGAACAGACACATTCCTCTGGT 609679

QY 1048 TTCACACACCGGGATATCAGCAGTGGGGCGGCGCACGACTCGATGTTTCATCGCCGAGTC 1107
DB 609680 TATTCTTACGAAATAATGCCAAGTGGTGCAGGACATGATGCAATGTCATATGGCAACACTT 609739

QY 1108 ACGGAGCTGCGAATGGTTTTCGTTTCCAAAGCCGCTGCGCGGAGCCAGGTTCCCGAAGAA 1167
DB 609740 TGCCCAACAGATATGATTTTATTTCCATCTCATTTGGGAATTAGTCAATACCCCTCTTGA 609799

QY 1168 TGGACCGATTTGATGACCTTCGCAAAAGGAACTGAGGTGTGCTCCGCGGTAAATGAAG 1225
DB 609800 TTTACTGATGGAAGATATAGAGCGGAGGAATTAAGTTTTTACAAAAAGTTATATATGG 609857
```

RESULT 14
US-10-329-670-1
; Sequence 1, Application US/10329670

```
Publication No. US20040018503A1
GENERAL INFORMATION:
APPLICANT: Fleischmann et al.
TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genome, Frag
FILE REFERENCE: PB186P1
CURRENT APPLICATION NUMBER: US/10/329,670
CURRENT FILING DATE: 2002-12-24
PRIOR APPLICATION NUMBER: US 09/643,990
PRIOR FILING DATE: 2000-08-23
PRIOR APPLICATION NUMBER: US 08/487,429
PRIOR FILING DATE: 1995-06-07
PRIOR APPLICATION NUMBER: US 08/426,787
PRIOR FILING DATE: 1995-04-21
NUMBER OF SEQ ID NOS: 1
SOFTWARE: PatentIn version 3.1
SEQ ID NO 1
LENGTH: 1830121
TYPE: DNA
ORGANISM: Haemophilus influenzae
FEATURE:
NAME/KEY: misc feature
LOCATION: (4747)..(4747)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (9921)..(9921)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (10150)..(10150)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (29298)..(29298)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36543)..(36543)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36551)..(36551)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (36636)..(36636)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (40808)..(40810)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44416)..(44416)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44905)..(44905)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (44975)..(44975)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45593)..(45593)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (45732)..(45732)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (47036)..(47036)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51334)..(51334)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51602)..(51602)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51786)..(51786)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (51805)..(51805)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (55369)..(55369)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65309)..(65309)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (65313)..(65313)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (80024)..(80024)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (100091)..(100091)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (102696)..(102696)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (105121)..(105121)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (107248)..(107248)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (117136)..(117136)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119750)..(119750)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (119924)..(119924)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (120038)..(120038)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
LOCATION: (121344)..(121344)
OTHER INFORMATION: n equals a, t, g or c
FEATURE:
NAME/KEY: misc feature
```

```
/ LOCATION: (122167)..(122167)
/ FEATURE:
/ OTHER INFORMATION: n equals a, t, g or c
/ NAME/KEY: misc_feature
/ LOCATION: (122336)..(122336)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131340)..(131340)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (131360)..(131360)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (139910)..(139910)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (140398)..(140398)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (142750)..(142750)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145171)..(145171)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (145942)..(145942)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (147197)..(147197)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (150841)..(150841)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (152500)..(152500)
/ OTHER INFORMATION: n equals a, t, g or c
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (152530)..(152530)
```

```
Query Match          9.8%; Score 121.2; DB 16; Length 1830121;
Best Local Similarity 43.8%; Pred. No. 2.7e-25;
Matches 525; Conservative 0; Mismatches 673; Indels 0; Gaps 0;

QY 28 CGCATGAGAAAGAGATCCGGAGCTCTCCCGGTTCTCGGCAGAGGCCCGGTTGTTACC 87
Db 608660 CGTGTTCAAATTTAAATTTGAAATACTGGCTTTTATTTTCATCAGTACCAAAACGAGCTGACT 608719

QY 88 CGGCTGACCTACACTCCAGAGATCCGCCGCGGGAACCTCATTTGCGGCTATGAA 147
Db 608720 CGCTTAGCTTTTACAGAGAGATGAAAGGCCCAATATGATTATTGAAATATTGTA 608779

QY 148 GCGGCGCCTTGAGGCTTCGTGAAGACGCACTCGGAAACATCATCGCGCAGCTGAAGGC 207
Db 608780 GAATATGATTTCATTCATTCGTGAGATTCATTTGGAATCTTTTATTTCGTAAGCAGGT 608839

QY 208 ACTGATCCGGAGCTTCCTCGGATCCGGGTCGGTTCACATTCGATTCGTTCGGAACGGC 267
Db 608840 AAAGAAGATTTTTCCTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCG 608899
```

```
QY 268 GGGATGTTTGATGGCACTGCGAGCGTGTGTGCGCCCTTGAGGCTGCGCGGTGATGCTG 327
Db 608900 GGTAAATTTGATGGTCCTTTGGGATCTGTTGCTGGTTGGAATTTCTTTTGAATTTGCT 608959

QY 328 GAGAACGGCTACGTGAATCGGCATCCATTTGAGTTTCATCGCATCGTGGAGGAGAAAGG 387
Db 608960 GAAACAGAAATATTCAGACTCGTTATCCTTTAGAAATGATAATTTTACTTGTGAAGAAATCG 609019

QY 388 GCCCGCTTCAGCAGTGGCATGTTGGGCGCGCGGCCATTCAGGGTTTGGTCCGCCACAGG 447
Db 609020 AGTAGATTTAAATTTTGTACATTCGGGTAGTAAAGTTATGTGGCATAGTAAATCAAGAA 609079

QY 448 GAACTGGACTCTTTGGTTGATGAGATGGAGTGTCCGTTAGGCGAGCGGCTACTCCCTTC 507
Db 609080 AAATTAAGTTCAATACGTGATAAACAAGAAAGGGTTTATCAGAAGCTATGGCTGAAGTA 609139

QY 508 GCGTTGAAGCCGGCGGAACCTGAGGCTGCAGCCCGCTCCGGCGGAGACTCGGTGCTTTT 567
Db 609140 GGAATGAATTTTAAATTTGTTTAAATCAAGCAAAACGTGATGCAAGGAATTTTAAATGTTT 609199

QY 568 ATCGAACTACACATTTGAACAGGACCGATCCTCGAGCAGGAGCAATAGAGATCGAGTT 627
Db 609200 TTTGAACCTTCATAGAAACAAGGCCCTCGTTTAGAGATGAAGGAAACCAATAGGTGTT 609259

QY 628 GTAACCTCCATCGTTTGGCGTTTCGCGCATTTGCGGTTGCGGTCAAAGGCGAGAACGCCAC 687
Db 609260 GTGACAGGTATTTGTCGCCCGATTCGTGCAATTTGTTAAATTAAGGACAGCGGATCAT 609319

QY 688 GCCGCAACAACCCCATGCACTCGCCAGGATCGGTGTTACCCCGCTCTCATGTTG 747
Db 609320 TCAGGAGCAACAGCAATGCAATTCATGTCATGATGCTTTGTTAGGAGGATCTGAGTTATCA 609379

QY 748 AGGAGGTCAACCGGTTGCTCAACAGATCCCGATGSCACAGTGGCTTACCGTTGGCCAC 807
Db 609380 CTTGCTATTGAGCGAGCTGCTATTCAAGCTGGACATTCACAGTAGCTACAGTAGGTAAAT 609439

QY 808 CTCACAGTGGCCCGCGGTGGAGCAACCCAGTCCCGGGGAGGTGGACTTCACACTGGAC 867
Db 609440 ATTACAGCTAAACAGGNGTAAATGAAATTTGTGCCAGATATTGCGAATTTATTAGTAGAT 609499

QY 868 CTGCTTCTCGCATGAGGAGTCGCTCCGCTGCTGATGATGATGATTTTCAATCTCAGC 927
Db 609500 ATTCAGGTACACATGTACAAAGCTAGAGATCTGTATTGAAATATTATCAAGAGGAAAT 609559

QY 928 GCGAGGTGCGCTCCAGGCCGCTGCTGCTCCGATGATGATGATTTTCAATCTCAGC 987
Db 609560 AGTAAAGTTTCAGAAAAAAGAGGATTTAATAGTTTACAACTTTTTCAAAAGATAAAT 609619

QY 988 CCGGTGACGTGGCTTCCTTACCATGTTGGACCGCTTCGCGAAGCGGCTCGGCCCTTGCAG 1047
Db 609620 CCAATATATTACTGAAATATGTTAAACCAATAGCTGAAACAGCACATTCCTTGGT 609679

QY 1048 TTCACACACCGGATATCAGCAGTGGGCGGCGCAACGATCGATGTTTTCATCGCCAGGTC 1107
Db 609680 TATTTTACGAAATAATGTCAGGAGTGGTGCAGGACATGATGCAATGCAATGCAACACTT 609739

QY 1108 ACGACGTCGAAATGGTTTTCGTTCCAAAGCGTGTGCGCGGAGCCAGTTTCCCGAAGAA 1167
Db 609740 TGCCCAACAGGTATGATTTTATTTTTCATCTCATTTTGGGAATTTAGTCATAACCTCTTCAA 609799

QY 1168 TGGACCGATTTTCGATGACTTCGCAAAAGGAACCTGAGGTTGCTCCCGGGTAAATGAAG 1225
Db 609800 TTTACTGATTGGAAGATATAGAGCAGGAAATTAAGTTTTTACAAAAAGTTATATTGG 609857
```

```
RESULT 15
US-10-156-761-3482
; Sequence 3482, Application US/10156761
; Publication No. US200301190181
; GENERAL INFORMATION:
; APPLICANT: OMURA, SATOSHI
; APPLICANT: IKEDA, HARUO
```


This Page Blank (uspto)